THE CRYSTAL STRUCTURES OF

EPI-LIMONOL IODOACETATE

AND AZULENE

THESIS

PRESENTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE

UNIVERSITY OF GLASGOW

BY

DAVID G. WATSON, B.Sc.

CHEMISTRY DEPARTMENT

DECEMBER, 1959

PREFACE

This thesis describes the determination, by the methods of X-ray diffraction, of the crystal structures of epi-limonol iodoacetate and azulene. In the introductory chapter some of the methods used in this work are discussed. Details are given in Appendix 3 of some programmes devised for X-ray crystallographic calculations for use with English Electric 'DEUCE'.

I wish to acknowledge my thanks to my supervisors, Professor J.M. Robertson, Dr G.A. Sim and Dr H.M.M. Shearer for the advice and criticism which they have offered throughout the course of this work.

I am grateful also to Dr D.C. Gilles and the computing staff of the University of Glasgow for their help with computational problems and, in particular, to Mr D.G. Williams for his brief synopsis of 'DEUCE' and the available programming systems included in Appendix 3.1. To Professor R. Pepinsky and Dr V. Vand of Pennsylvania State University, Dr O.S. Mills of Manchester University, Dr R. Sparks recently at Oxford Computing Laboratory and Mrs Peters of the National Physical Laboratory I must also record my thanks for computations performed in their laboratories.

I wish to express my gratitude to Professor D.H.R. Barton for the supply of crystals of epi-limonol iodoacetate and to Mr Findlay of this department who has kindly prepared the photographs. To Miss M. Perry of the Computing Laboratory I am also grateful for

the typing of this thesis.

I am indebted to the Carnegie Trust for the Universities of Scotland for a research scholarship, and later to the University of Glasgow for appointment to an Assistant Lecturership, which enabled me to undertake the work.

It must be recorded that, because of the magnitude of the work, the analysis of epi-limonol iodoacetate has been a team project and the distribution of labour may be summarised thus:-

Unit Cell Parameters and Space Group	D.G. Watson and S. Arnott
Analysis of (100) Projection	D.G. Watson
Analysis of (010) Projection	S. Arnott
Recording and Measurement of Intensities	D.G. Watson, S. Arnott and A.W. Davie
Scaling of Data and Preparation of Punched Cards	D.G. Watson and A.W. Davie
Harker Section	D.G. Watson
Re-examination of (010) Projection	S. Arnott
3-D Patterson Synthesis Minimum Function 3-D Fourier Analysis Refinement of Structure	D.G. Watson, S. Arnott, A.W. Davie

At many stages throughout this work I have had consultations with my collaborators and supervisors as to the tactics to be employed.

SUMMARY

The crystal structures of two organic molecules have been determined by the methods of X-ray diffraction.

Epi-limonol iodoacetate, $C_{28}H_{33}O_9I$, belongs to the monoclinic space group $P2_1$ with four molecules per unit cell. The crystal structure analysis has been based on the heavy-atom technique. Three-dimensional data have been used for the computations, the chief methods used being the Patterson synthesis and corresponding minimum function combined with Fourier syntheses. This analysis has led to a complete determination of the molecular structure and stereochemistry of epi-limonol iodoacetate and hence of the parent compound, limonin, $C_{26}H_{30}O_8$. The latter is of great interest to organic chemists and the determination of its structure may make a significant contribution to the elucidation of other closely-related structures.

Azulene, $C_{10}H_8$, is monoclinic, space group $P2_1/a$, with two molecules per unit cell. The partially refined structure, postulated as belonging to the space group Pa, has been analysed by the method of least-squares and found to be incapable of refinement. Re-examination of the space group requirements has led to a statistically disordered structure in the space group $P2_1/a$. This structure has been exhaustively refined by three-dimensional least-squares analysis but the results are rather unsatisfactory. The extreme 'closeness' of pairs of atoms prohibits a very accurate

determination of atomic coordinates. However, the transannular C-C bond has been shown to be considerably larger than the other bonds in the molecule in agreement with theoretical considerations.

Several programmes have been developed for use with English Electric 'DEUCE'. These programmes, written with interpretive coding schemes, facilitate various computations encountered in X-ray crystallographic analysis.

C O N T E N T S

CHAPTER I

SOME METHODS OF CRYSTAL STRUCTURE ANALYSIS

1.1 Structure Factor Formulae

For any reflection (hkℓ) the structure factor is a complex function whose modulus, the structure amplitude, is defined as the ratio of the amplitude of the radiation scattered in the order h, k, ℓ by the contents of one unit cell to that scattered by a single electron under the same conditions (Lonsdale, 1936). The structure factor may be expressed analytically by

$$F(hk\ell) = \sum_{j=1}^{N} f_j \exp \left\{ 2\pi i(hx_j + ky_j + \ell z_j) \right\} \qquad \ldots\ldots\ldots\ldots(1)$$

where f_j are the scattering factors of the atoms j, N is the total number of atoms in the unit cell and x_j, y_j, z_j, are the atomic coordinates expressed as fractions of the cell edges. The atomic scattering factor, f_j, is a function of the scattering angle and depends on the distribution of electrons in the atom. These distributions have been calculated for various atomic species by, among others, Hartree (1928), James and Brindley (1932), McWeeny (1951, 1952) and Hoerni and Ibers (1954). The atomic scattering factor is calculated for an atom at rest, but thermal vibrations tend to make the electron distribution more diffuse and decrease the scattering power.

If f_0 represents the scattering factor for an atom at rest then the true scattering factor is given by

$$f = f_0 \exp \left\{ -B(\sin \theta/\lambda)^2 \right\} \qquad \ldots\ldots\ldots\ldots(2)$$

The constant, B, the temperature factor, is related to the mean square displacement, $\overline{u^2}$, of the atoms from their mean positions by the expression

$$B = 8\pi^2 \overline{u^2} \qquad \ldots\ldots\ldots\ldots\ldots(3)$$

The structure factor equations for the 230 space groups have been conveniently listed in International Tables (1952).

1.2 Representation of Electron Density by Fourier Series

It was first suggested by Bragg (1915) that since a crystal consists of an infinitely repeating array of unit cells the electron density could be represented by a triple Fourier series. Thus the electron density $\rho(xyz)$ at the point (x, y, z) can be represented by the equation

$$\rho(xyz) = \sum_{-\infty}^{\infty}{}_{h'} \sum_{-\infty}^{\infty}{}_{k'} \sum_{-\infty}^{\infty}{}_{\ell'} C(h'k'\ell')\exp\left\{2\pi i(h'x+k'y+\ell'z)\right\} \qquad \ldots\ldots\ldots(4)$$

Now it can be shown that $C(\bar{h}\,\bar{k}\,\bar{\ell}) = F(hk\ell)/V$

$$\text{Thus } \rho(xyz) = \frac{1}{V}\sum_{-\infty}^{\infty}{}_{h} \sum_{-\infty}^{\infty}{}_{k} \sum_{-\infty}^{\infty}{}_{\ell} F(hk\ell)\exp\left\{-2\pi i(hx+ky+\ell z)\right\} \qquad \ldots\ldots\ldots(5)$$

Since the atomic scattering factors, f_j, decrease with increasing $\sin\theta/\lambda$, the coefficients, $F(hk\ell)$, in the electron density function fall off and the Fourier series is convergent. It is

convenient to write (5) in the form

$$\rho(xyz) = \frac{1}{V} \sum_{-\infty}^{\infty}_{h} \sum_{-\infty}^{\infty}_{k} \sum_{-\infty}^{\infty}_{\ell} |F(hk\ell)| \cos\left\{2\pi(hx+ky+\ell z)-\alpha(hk\ell)\right\} \quad \ldots\ldots\ldots(6)$$

where $\alpha(hk\ell)$ is the phase constant associated with the amplitude $|F(hk\ell)|$.

Because of the great amount of computational labour involved, triple Fourier series are generally only used where very high accuracy is sought, or in the case of very complex molecules.

In many structure analyses, it is sufficient to consider projections of the electron density distribution on to planes.

Thus $$\rho(xy) = \frac{1}{A} \sum_{-\infty}^{\infty}_{h} \sum_{-\infty}^{\infty}_{k} |F(hk0)| \cos\left\{2\pi(hx+ky)-\alpha(hk0)\right\} \quad \ldots\ldots\ldots(7)$$

represents the electron density projected on to the (x, y) plane. This double Fourier series method was first employed by Bragg (1929) in the solution of the diopside structure and has proved of great use in the case of planar aromatic hydrocarbons (Robertson, 1953). By combining the results from two or more projections the complete structure can often be established with very great accuracy.

2.1 The Phase Problem

Thefundamental problem of X-ray crystallography lies in the fact that, although the observed intensities of the X-ray reflections can be used to calculate $|F(hk\ell)|$, the corresponding phase

constants, $\alpha(hk\ell)$, cannot be measured. Accordingly, an infinite number of different electron density distributions may be derived from the $|F_0|$ values by assigning arbitrary phase constants to each structure amplitude and summing the resulting Fourier series.

However, although there is no unique mathematical solution to the phase problem, it may be possible to find a unique physical solution by consideration of the following criteria:-

a) the electron density should be everywhere positive

b) the atoms should be approximately spherically symmetrical

c) the structure should be chemically reasonable.

Many attempts have been made to solve the phase problem, with greater success in the case of centrosymmetrical structures since in such systems the phase angles are restricted to 0 or π. This corresponds to positive or negative signs of the coefficients of the Fourier series.

$$\text{Thus} \quad \rho(xy) = \frac{1}{A} \sum_{-\infty}^{\infty h} \sum_{-\infty}^{\infty k} \pm F(hk0) \cos\left\{2\pi(hx+ky)\right\} \quad .....(8)$$

The methods employed in the work reported in this thesis will now be outlined.

2.2 Trial-and-Error Method

By consideration of the chemical structure, if known, and

the more intense X-ray spectra a trial set of atomic coordinates can often be postulated for which structure amplitudes may be calculated. If the latter show reasonable agreement with the observed values then the calculated phase constants together with the observed amplitudes may be used to evaluate a Fourier series which should give better atomic positions. Iterations of this procedure can then lead to a fairly accurate structure determination.

2.3 The Patterson Function

An attempt to overcome the phase problem was made by Patterson (1935) using the squares of the moduli of the structure factors as Fourier coefficients. These quantities are directly related to the observed intensities and can always be measured. The Patterson function, P(uvw), is defined by the expression

$$P(uvw) = V \int_0^1 \int_0^1 \int_0^1 \rho(xyz)\, \rho(x+u, y+v, z+w)\, dx\,dy\,dz \quad \ldots\ldots(9)$$

This may be reduced, by integration, to the form

$$P(uvw) = \frac{1}{V} \sum_{-\infty}^{\infty}{}_h \sum_{-\infty}^{\infty}{}_k \sum_{-\infty}^{\infty}{}_\ell \left| F(hk\ell) \right|^2 \exp\left\{-2\pi i(hu+kv+\ell w)\right\} \quad \ldots\ldots(10)$$

Collecting together the coefficients in pairs, $hk\ell$ and $\bar{h}\,\bar{k}\,\bar{\ell}$, (10) may be written

$$P(uvw) = \frac{1}{V} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} |F(hk\ell)|^2 \cos\{2\pi(hu+kv+\ell w)\} \quad \ldots\ldots (11)$$

which is real for all values of u, v, w.

The physical meaning of P(uvw) can be understood by a consideration of (9). P(uvw) can clearly only have large values when both $\rho(xyz)$ and $\rho(x+u, y+v, z+w)$ are large. This occurs if there are atoms at (x, y, z) and (x+u, y+v, z+w) separated by the vector distance (uvw). Thus a peak in the Patterson function at (u_1, v_1, w_1) corresponds to an interatomic distance in the crystal defined by a vector whose components are u_1, v_1, w_1.

The interpretation of Patterson maps is generally fairly difficult, especially in projection, for the following reasons:-

a) for a system of n atoms there will be n(n-1)/2 distinct peaks in the vector distribution and these will tend to overlap.

b) the theory is strictly valid only for point atoms and the peaks will tend therefore to be broad and ill-defined.

Various methods have been studied in an attempt to improve the resolution, mainly by Patterson (1935) and Yu (1942).

However, if there is present in the unit cell a small number of heavy atoms there is a strong possibility of interpreting the Patterson function. In this case the vectors between the heavy atoms, and, perhaps between the heavy and light atoms, give rise to peaks which stand out against the background of peaks due to

the light atoms.

2.4 The Harker Synthesis

Harker (1936) has developed a useful modification of the Patterson function involving the symmetry properties of the crystal. Consider a crystal with a 2-fold screw axis which will be taken as coincident with the b-axis. The equivalent positions are represented by (x, y, z) and $(\bar{x}, \frac{1}{2}+y, \bar{z})$. The vector between atoms at these positions has components $(2x, -\frac{1}{2}, 2z)$ and there will be a maximum in P(uvw) at the point $(2u, -\frac{1}{2}, 2w)$. This maximum lies in the plane $v = -\frac{1}{2}$, or $v = \frac{1}{2}$, since P(uvw) is centrosymmetric, and there will be a maximum in the plane $v = \frac{1}{2}$ for each crystallographically different kind of atom in the crystal. (11) may then be written in the form

$$P(u,\tfrac{1}{2},w) = \frac{1}{V} \sum_{-\infty}^{\infty}{}_{h} \sum_{-\infty}^{\infty}{}_{l} C_{hl} \cos\left\{2\pi\,(hu+lw)\right\} \quad \ldots\ldots(12)$$

where $$C_{hl} = \sum_{-\infty}^{\infty}{}_{k} (-1)^{k} \left| F(hkl) \right|^{2}$$

(12) is effectively only a two-dimensional summation and there is a good chance of resolving the vector peaks for

a) all the measured $|F(hkl)|$ values are used

b) interatomic vectors not parallel to the plane $v = \frac{1}{2}$ are eliminated.

One difficulty lies in the fact that certain atoms not related

by the symmetry elements may have one or two coordinates which are identical and result in non-Harker peaks.

2.5 The Heavy-Atom Technique

The heavy-atom technique is based on the presence of one or more atoms in or associated with the molecule, these atoms being of predominantly higher scattering power than the remaining atoms. The positions of these heavy atoms can be located fairly easily by the Patterson method as indicated above, and combination of the observed structure amplitudes with phase angles calculated from the parameters of the heavy atoms alone provides coefficients for a Fourier synthesis which is a close approximation to the complete structure. This initial Fourier summation will generally show a number of small peaks, some genuine and some spurious. Further Fourier refinement may remove the spurious peaks or, by a consideration of accepted bond lengths and bond angles, peaks corresponding to real atoms may be distinguished. The result can be expressed analytically by writing the expression for the structure factor in the case of a crystal with one heavy atom in the unit cell:-

$$F(hk\ell) = f_H \exp\left\{2\pi i(hx_H+ky_H+\ell z_H)\right\} + \sum_{j=1}^{n} f_j \exp\left\{2\pi i(hx_j+ky_j+\ell z_j)\right\}$$

.....(13)

where f_H is the scattering factor of the heavy atom, whose parameters are x_H, y_H, z_H and $\underline{n}$ is the number of "light" atoms. It is inadvisable that f_H should be very much greater than $\sum_{j=1}^{n} f_j$ since the resulting Fourier synthesis would tend to show only the heavy atom. A rough rule for application to this method is that the sum of the squares of the atomic numbers of the heavy atoms and of the light atoms should be approximately equal.

It should be noted that this technique is capable of being independent of chemical information regarding the molecule and has proved successful in the elucidation of many structures of which may be mentioned cholesteryl iodide (Carlisle and Crowfoot, 1945), β-caryophyllene bromide (Robertson and Todd, 1953) and lanostenol (Curtis et al, 1952).

2.6. Method of Superposition and the Minimum Function

Suppose the structure has a centre of symmetry and one heavy atom per asymmetric unit. The coordinates of this atom may be determined from the Patterson function, as, say, (x_1, y_1, z_1). In the method developed by Beevers and Robertson (1950) the Patterson function is then placed with its origin at each of the two known atomic positions (x_1, y_1, z_1) and $(\bar{x}_1, \bar{y}_1, \bar{z}_1)$ in turn, to give a composite Patterson function

$$P_2(xyz) = P(x-x_1, y-y_1, z-z_1) + P(x+x_1, y+y_1, z+z_1) \quad .....(14)$$

In this function the images of other atoms in each of the heavy atoms are brought into coincidence on atomic sites. Whether the latter can be recognised in practice depends on the relative weights of the atoms involved.

In the general case of a non-centrosymmetrical structure in which the positions of $\underline{m}$ atoms are known, a superposition of $\underline{m}$ identical Patterson functions with origins at the points (x_1, y_1, z_1), (x_2, y_2, z_2) ..... (x_m, y_m, z_m) results, on addition, in a function

$$P_m(xyz) = \sum_{j=1}^{m} P(x-x_j, y-y_j, z-z_j) \qquad \text{.....(15)}$$

Buerger (1951) has shown that there are certain advantages in forming, instead of $P_m(xyz)$, the minimum functions $M_m(xyz)$ which is the lowest value of $P(x-x_j, y-y_j, z-z_j)$ for any value of j. There is a high probability that peaks in the functions $M_m(xyz)$, especially if $\underline{m}$ is fairly large, will represent true atomic sites.

3.1 Methods of Refinement

Various methods have been devised for the refinement of positional and thermal parameters in an effort to make accurate determinations of molecular geometry and vibrations of molecules in crystal lattices. Among the more important may be mentioned the F_0 synthesis, the differential synthesis (Booth, 1946),

the method of steepest descents (Booth, 1947, 1949; Vand, 1948, 1949, 1951; Qurashi, 1949), the (F_o-F_c) synthesis (Brindley and Wood, 1929; Crowfoot et al. 1949; Booth, 1948; Cochran, 1951) and the method of least squares.

3.2 Method of Least Squares

Least squares analysis was first applied to crystal structure refinement by Hughes (1941) and has since been used successfully by many workers. In this method F_o and F_c are compared and the problem is to find the atomic parameters which will render the sum of the squares of the discrepancies a minimum, according to Legendre's principle. Thus the quantity to be minimised is given by

$$R' = \sum \sqrt{w} \left(|F_o| - |F_c| \right)^2 \quad \text{.....(16)}$$

Since the errors in F_o may vary to some extent, weighting factors must be applied and should be inversely proportional to the square of the probable error of the corresponding F_o. The least squares routine is based on the assumption that the shifts to be applied are small enough to be treated as differentials. For every observed spectrum one can set up an observational equation of the type

$$\sum_{i=1}^{N} \left(\sqrt{w} \frac{\partial |F_c|}{\partial \xi_i} \right) \Delta \xi_i = \sqrt{w} \left(|F_o| - |F_c| \right)^2 \quad \text{.....(17)}$$

where $\Delta\xi_i$ are the corrections to be solved for and applied to the values of ξ_i used in calculating the $|F_c|$ values. These observational equations are then normalised by the least squares procedure. Reduction of the computational labour is often achieved by making the following assumptions:-

a) if the atoms are well-resolved, quantities $\sum\sqrt{w}\left(\frac{\partial|F_c|}{\partial\xi_i}\cdot\frac{\partial|F_c|}{\partial\xi_j}\right)$ will be small compared with $\sum\sqrt{w}\left(\frac{\partial|F_c|}{\partial\xi_i}\right)^2$

b) if the axes are orthogonal, quantities $\sum\sqrt{w}\left(\frac{\partial|F_c|}{\partial\xi_i}\cdot\frac{\partial|F_c|}{\partial\xi'_i}\right)$ can be neglected.

Thus considering only the diagonal elements of the matrix of normal equations one can represent the normal equations by

$$\Delta\xi_i\sum_{i=1}^{N}\sqrt{w}\left(\frac{\partial|F_c|}{\partial\xi_i}\right)^2=\sum_{hk\ell}\sqrt{w}(|F_o|-|F_c|)\frac{\partial|F_c|}{\partial\xi_i}\qquad\ldots\ldots(18)$$

When refinement is continued until R and $\sum\sqrt{w}\,|\Delta F|^2$ approach constancy the parameter shifts - positional, vibrational and scale - indicated at that stage should be only a small fraction of the corresponding standard deviations.

3.3 Agreement Index

A useful numerical test of the agreement between F_o and F_c

values which is generally adopted is the agreement index

$$R = \sum |F_o - F_c| \; / \; \sum |F_o| \qquad \ldots\ldots\ldots\ldots\ldots\ldots (19)$$

Expressed as a percentage, 100R is the percentage discrepancy. Although R does not correspond to the function minimised by the Fourier or least squares method it does provide a reasonable indication of the progress of refinement.

CHAPTER II

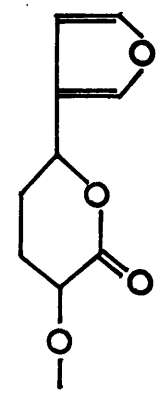
THE CRYSTAL STRUCTURE OF EPI-LIMONOL IODOACETATE

AND MOLECULAR STRUCTURE OF LIMONIN

1. Introduction

Limonin, $C_{26}H_{30}O_8$, the bitter principle of citrus fruits, was first isolated in 1841 (Bernays), but has only been studied intensively over the past ten to twenty years. Shortly after X-ray studies were initiated on limonin and its derivatives some information became available (Melera et al, 1957) concerning the known chemical features of the parent molecule and can be summarised thus:-

Limonin contains two lactone rings, which can be opened reversibly, a β-substituted furan ring, a ketonic oxygen atom and two ethereal oxygen rings. Infra-red measurements reveal that both lactone rings are δ-constituted and a study of the hydrogenolysis products indicates that the allyl oxygen of the lactone which is cleaved is attached allylicly with respect to the furan ring. This situation may be compared to that which exists in columbin (Barton and Elad, 1956), the main bitter principle of Colombo root. Hexahydrolimoninic acid, $C_{26}H_{36}O_8$, formed by hydrogenation of limonin, is an abnormally strong acid and it was postulated (Melera et al, 1957) that this could be explained by an oxygen substituent α to the carboxyl group. The ketone group of limonin has an I.R. frequency corresponding to that of a six-membered (or higher) ketone. At this stage it was possible to assign the part formula I.



I

The molecular formula of limonin in conjunction with the established functional groups requires that the molecule be bicarbocyclic and since vigorous degradation of limonin (Koller and Czerny, 1936; Brachvogel, 1952) yields 1:2:5-trimethylnaphthalene it can be reasonably supposed that the two carbocyclic rings are both six-membered.

2. Previous X-Ray Examination of Limonin and its Derivatives

A survey of the unit cell parameters and space groups of limonin and several derivatives has been conducted by Arnott and Robertson (1959) and their results for limonin and its acetic acid solvate are similar to those reported by Jones and Palmer (1949). Because of the size and complexity of the molecule it was decided that the best approach would be by means of some phase-determining heavy-atom technique (Robertson, 1935, 1936; Robertson and Woodward, 1937, 1940). Accordingly, attention was focussed on

Table 1

Crystallographic data for epi-limonol iodoacetate

and epi-limonol chloroacetate

	epi-limonol iodoacetate	epi-limonol chloroacetate
Molecular formula	$C_{26}H_{31}O_8(COCH_2I)$	$C_{26}H_{31}O_8(COCH_2Cl).H_2O$
a(Å)	15·03 ± 0·02	12·26 ± 0·02
b(Å)	12·36 ± 0·02	10·92 ± 0·02
c(Å)	15·93 ± 0·02	11·50 ± 0·02
β	95°12' ± 15'	93°15' ± 15'
U(Å^3)	2952·7	1836·4
Z	4	2
D_m(g.cm^{-3})	1·426	1·220
D_x(g.cm^{-3})	1·441	1·228
μ (cm^{-1}), λ = 0·7107 Å	12·3	2·0
Mol. wt.	640·5	568·0
Space Group	$P2_1(C_2^2)$	$P2_1(C_2^2)$

the two esters, epi-limonol iodoacetate and epi-limonol chloroacetate.

3. Experimental Details

3.1. Preparation of the Crystals, Unit Cell Parameters and Space Groups

Crystals of epi-limonol iodoacetate and chloroacetate were grown from solutions in acetone-water, the former appearing as plates and the latter as thin needles.

The unit cell parameters were obtained from rotation and equatorial layer-line moving film photographs. Later more accurate cell dimensions were determined from precession photographs, marked with fiducial spots, taken with molybdenum Kα radiation (λ=0·7107 Å). The crystallographic data for the two esters are listed in Table 1. It should be noted that the systematic absences - (0k0) for k = (2n+1) - are characteristic of both $P2_1(C_2^2)$ and $P2_1/m(C_{2h}^2)$, but the latter space group was rejected as being incompatible with the fact that both esters are optically active.

3.2. Choice of Derivative for Structural Investigations

Since the iodoacetate and chloroacetate were found to be not isomorphous a choice had to be made for further investigations. Inspection of Table 1 reveals that the latter has a much smaller linear absorption coefficient for X-rays, (μ), and has a smaller unit cell containing only two chemical molecules related

by the screw-axis. However the choice of an origin mid-way between the two phase-determining chlorine atoms would have the disadvantage of introducing the ambiguity of a false symmetry centre in the subsequent analysis and thus phase angles calculated on the basis of the chlorine atoms alone would be 0 or π. Consideration of the various factors led to the conclusion that the iodoacetate offered a greater chance of success in spite of the following features:-

In the space group $P2_1(C_2^2)$ the general positions are two-fold, and hence, with four molecules in the unit cell, the asymmetric crystal unit consists of two chemical molecules. This is a considerable complication, because it means that the coordinates of 76 atoms other than hydrogen atoms must be determined from the X-ray data. However, as will be seen later, this situation has been of considerable use in confirming our results.

4. Two-Dimensional Analysis of Epi-limonol Iodoacetate

4.1. Recording of X-Ray Data, Measurement and Correction of Intensities

The X-ray data used in the two-dimensional survey were obtained from equatorial layer-line moving film photographs around the a- and b- axes, using a Weissenberg camera with $CuK\alpha$ radiation. Robertson's multiple-film technique (1943) was employed

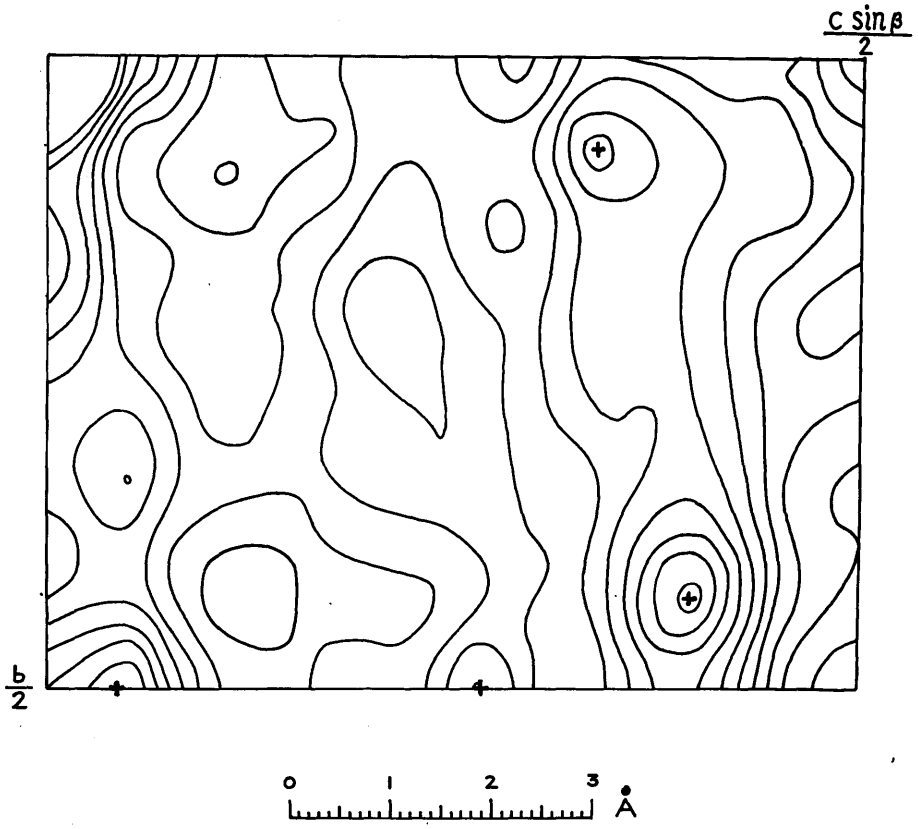


Fig. 1 Projection of Patterson function on (100).

for the correlation of strong and weak reflections and the intensities estimated visually. Corrections for Lorentz and polarisation factors were made assuming the usual mosaic crystal formula:-

$$F^2 = I \sin 2\theta / (1 + \cos^2 2\theta) \qquad \ldots\ldots\ldots\ldots(20)$$

In this way a set of $|F_0|$ values on a relative scale was obtained.

4.2. Analysis of (100) Projection

For the successful application of the heavy-atom technique it was essential to determine the coordinates of the iodine atoms and this was most readily achieved by making use of the Patterson function.

Using the squares of the 158 observed structure amplitudes as Fourier coefficients a Patterson projection on (100) was calculated using Beevers-Lipson strips.

The resulting Patterson synthesis is shown in Fig. 1, the function having been evaluated from y = 0, to y = $\frac{1}{2}$ and from z = 0 to z = $\frac{1}{2}$. The y- and z- coordinates of the iodine atoms may be represented by

$$I_1 \quad (y_1, z_1) \qquad\qquad I_2 \quad (y_2, z_2)$$

$$I_1^* \quad (\tfrac{1}{2} + y_1, \bar{z}_1) \qquad\qquad I_2^* \quad (\tfrac{1}{2} + y_2, \bar{z}_2)$$

Thus the following vector peaks should occur:-

I_1-I_1^* $(\frac{1}{2}, 2z_1)$ $\quad$ I_1-I_2^* $(y_1-y_2-\frac{1}{2}, z_1+z_2)$

I_1-I_2 (y_2-y_1, z_2-z_1) $\quad$ I_2-I_2^* $(\frac{1}{2}, 2z_2)$

The vector peaks, denoted by crosses in Fig. 1, have maxima at (0·500, 0·049); (0·072, 0·338); (0·426, 0·398) and (0·500, 0·270). It was decided to choose the origin to be on the b-axis mid-way between the two iodine atoms. Making this choice the coordinates of the iodine atoms were determined to be

Atom	y	z
I_1	$\overline{0·037}$	0·025
I_2	0·037	0·365

Using these atomic coordinates the contributions of the iodine atoms to the (0k*l*) structure factors were calculated. The iodine scattering factor curve used was that due to Brindley and Wood (1931) with a temperature factor, B = 3·0 $Å^2$.

The phase angles calculated on the basis of the iodine atoms were associated with the corresponding observed structure amplitudes and an electron-density projection on (100) evaluated. The expression used for the calculation of the electron-density projection on (100) is

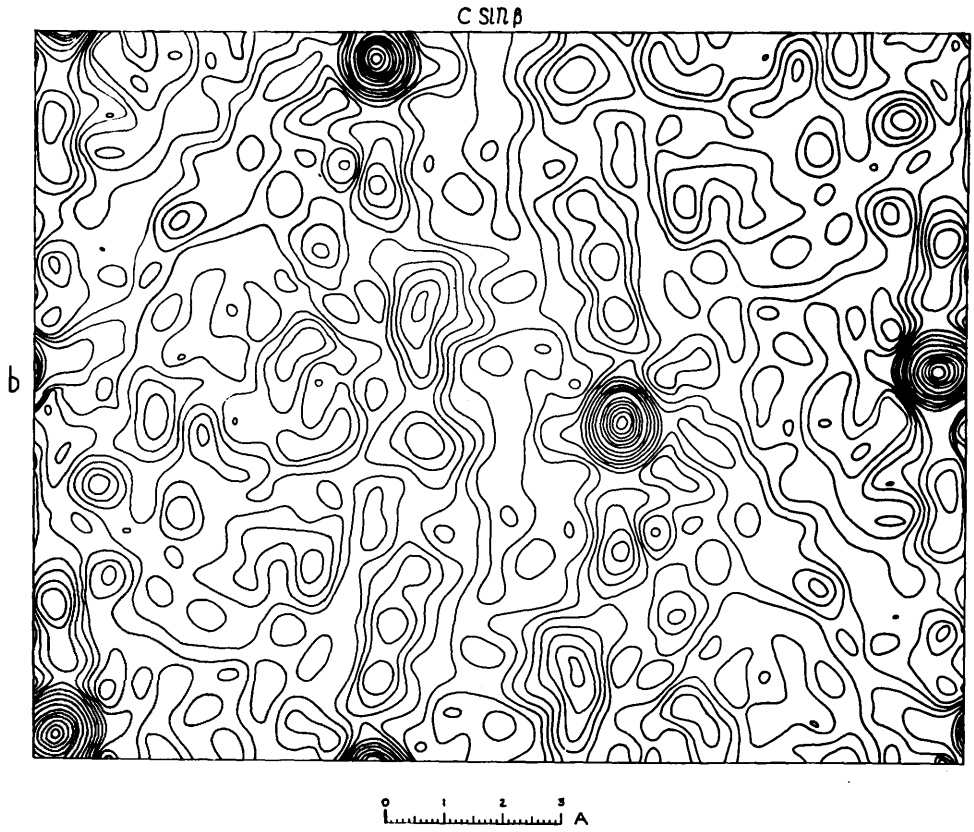


Fig. 2 Electron density distribution projected on (100).

$$\rho(YZ) = \frac{1}{A_c}\left[F(00) + 2\sum_{\ell=1}^{\infty} |F(0\ell)| \cos 2\pi\ell Z + \sum_{k=2}^{\infty}{}^{k=2n} |F(k0)| \right.$$

$$\cos\left[2\pi kY - \alpha(k0)\right]$$

$$+ 4\left\{ \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}{}^{k=2n} |F(k\ell)| \cos 2\pi\ell Z \cos\left[2\pi kY - \alpha(k\ell)\right] \right.$$

$$\left.\left. - \sum_{k=1}^{\infty}\sum_{\ell=1}^{\infty}{}^{k=2n+1} |F(k\ell)| \sin 2\pi\ell Z \sin\left[2\pi kY - \alpha(k\ell)\right] \right\} \right] \quad \ldots\ldots(21)$$

The resulting electron-density distribution, shown in Fig. 2, indicates quite clearly the iodine peaks but the remainder of the distribution is unresolved and cannot be interpreted in terms of molecular structure.

4.3. Analysis of (010) Projection

The Patterson function projected on (010) was computed using some 241 terms i.e. 38% of the total theoretical number of reflections for the (010) zone.

In this projection the heavy-atom positions may be designated by

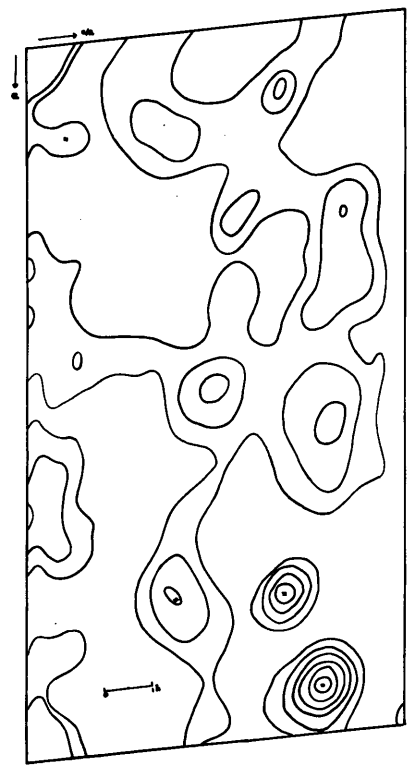


Fig. 3 Projection of Patterson function on (010).

I_1 (x_1, z_1) I_2 (x_2, z_2)

I_1^* $(\bar{x}_1, \bar{z}_1)$ I_2^* $(\bar{x}_2, \bar{z}_2)$

and the corresponding vector peaks should occur at

$I_1 - I_1^*$ $(2x_1, 2z_1)$ $I_1 - I_2^*$ (x_1+x_2, z_1+z_2)

$I_1 - I_2$ (x_1-x_2, z_1-z_2) $I_2 - I_2^*$ $(2x_2, 2z_2)$

The Patterson projection, illustrated in Fig. 3, shows peaks which enable the x- and z- coordinates of the iodine atoms to be determined.

Combination of the results obtained by a study of the Patterson projections on (100) and (010) yields the following coordinates for the iodine atoms:-

Atom	x	y	z
I_1	$\overline{0.062}$	$\overline{0.037}$	0·025
I_2	0·133	0·037	0·365

The x- and z- coordinates of the iodine atoms were used to calculate a set of structure factors for the (h0ℓ) data. An electron-density projection on (010) was calculated, using as

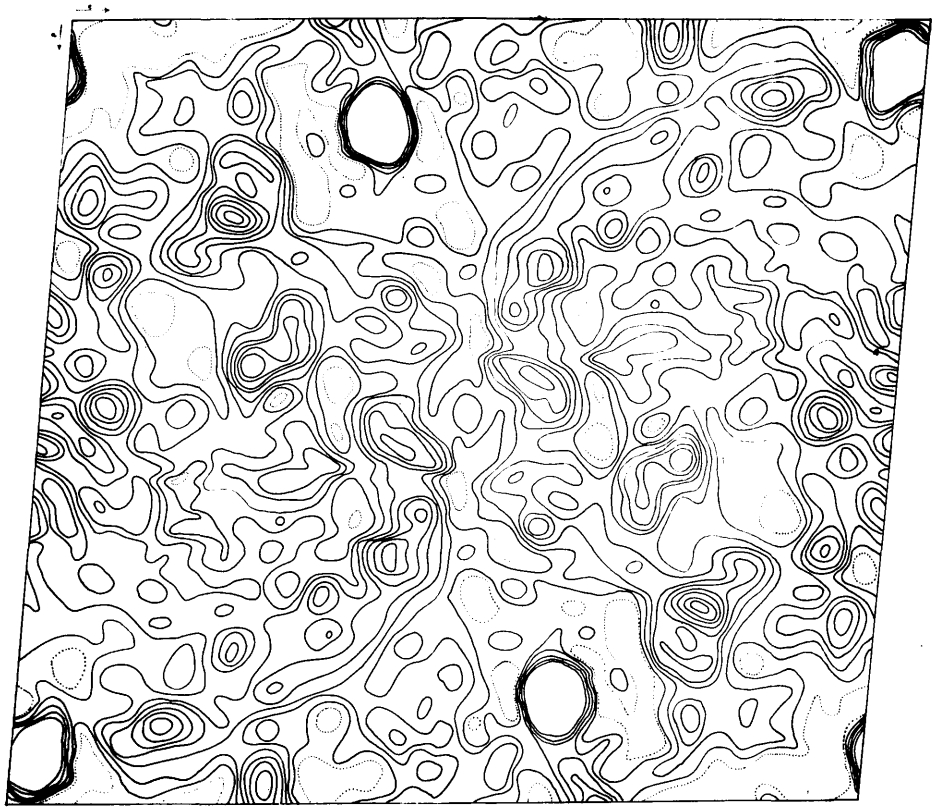


Fig. 4. Electron density distribution projected on (010). Contours above 25 electrons per $Å^2$ have been omitted.

Fourier coefficients the observed structure amplitudes with the signs of the structure factors calculated on the basis of the iodine atoms. The electron-density equation for the (010) projection is given by

$$\rho(XZ) = \frac{1}{A_c}\left[F(00) + 2\sum_{h=1}^{\infty} F(h0)\cos 2\pi hX + 2\sum_{l=1}^{\infty} F(0l)\cos 2\pi lZ + 2\sum_{h=1}^{\infty}\sum_{l=1}^{\infty}\left\{F(hl)\cos 2\pi(hX+lZ)+F(\bar{h}l)\cos 2\pi(-hX+lZ)\right\}\right] \quad \ldots\ldots\ldots\ldots\ldots(22)$$

At a later stage this electron-density computation was repeated using MoKα data and the resulting distribution is shown in Fig. 4. More terms were available for the Fourier synthesis in the case of the molybdenum data but the overall distribution is essentially the same as that obtained from the copper data. As with the (100) projection the only recognisable peaks are those due to the iodine atoms. However, the (010) electron-density distribution is not nearly so featureless as the (100) distribution and it will be of interest to reconsider the (010) projection later in the light of the complete structure. Nevertheless, it should be noted that with molecules of this complexity the possibility of solving the structure by consideration of projections of the electron-density is fairly small due to the very considerable overlap of atoms.

5. Three-Dimensional Analysis

5.1. Recording of X-Ray Data

For epi-limonol iodoacetate the linear absorption coefficients for X-rays have the values

$\mu = 96.0\ cm^{-1}$ for $CuK\alpha$ radiation ($\lambda = 1.542$ Å)

$\mu = 12.0\ cm^{-1}$ for $MoK\alpha$ radiation ($\lambda = 0.7107$ Å)

Accordingly, to minimise the introduction of errors due to absorption, $MoK\alpha$ X-radiation was used for all measurements. Photographic records were obtained using an equi-inclination Weissenberg camera and a precession camera. In the former case the multiple-film technique (Robertson, 1943), and in the latter case the multiple-exposure method, were used for the correlation of strong and weak reflections. In recording the Weissenberg photographs, sheets of nickel foil, 0·008 in. thick, were interleaved between films (Ilford 'Industrial-G'). Crystal specimens measuring about 1·0 x 0·2 x 0·02 mm. were used and it was found that the ratio of the strongest to the weakest intensity in any layer was about 8000: 1. The reciprocal lattice was explored by recording the intensities of the $(h0l)$-$(h8l)$ and $(0kl)$-$(6kl)$ layers with a Weissenberg camera and the $(hk0)$-$(hk3)$ layers with a precession camera.

5.2. Measurement and Correction of Intensities

The intensities were estimated visually by independent observers using the 'standard-spot' and 'step-wedge' techniques. Absorption corrections, being small, were not applied. Intensities measured on the precession camera films were corrected by the Lorentz and polarisation factors using the charts of Waser (1951) and Grenville-Wells and Abrahams (1952). Those measured on the Weissenberg films were corrected by the usual Lorentz and polarisation factors and by the Tunell (1939) rotation factor.

6. Structure Determination

6.1. Harker Section

As indicated in 2.4 of Chapter I, Harker (1936) has shown how, if one considers the symmetry elements of the crystal, the most useful information to be derived from a three-dimensional Patterson function is often concentrated over a particular plane or along a particular line. This approach reduces the computational labour to that of a two-dimensional summation but with the use of three-dimensional data.

Since the complete structure analysis of epi-limonol iodoacetate was to be based on the determined coordinates of the iodine atoms it was decided to compute the Patterson function for the appropriate Harker section in order to obtain as accurate iodine coordinates as possible.

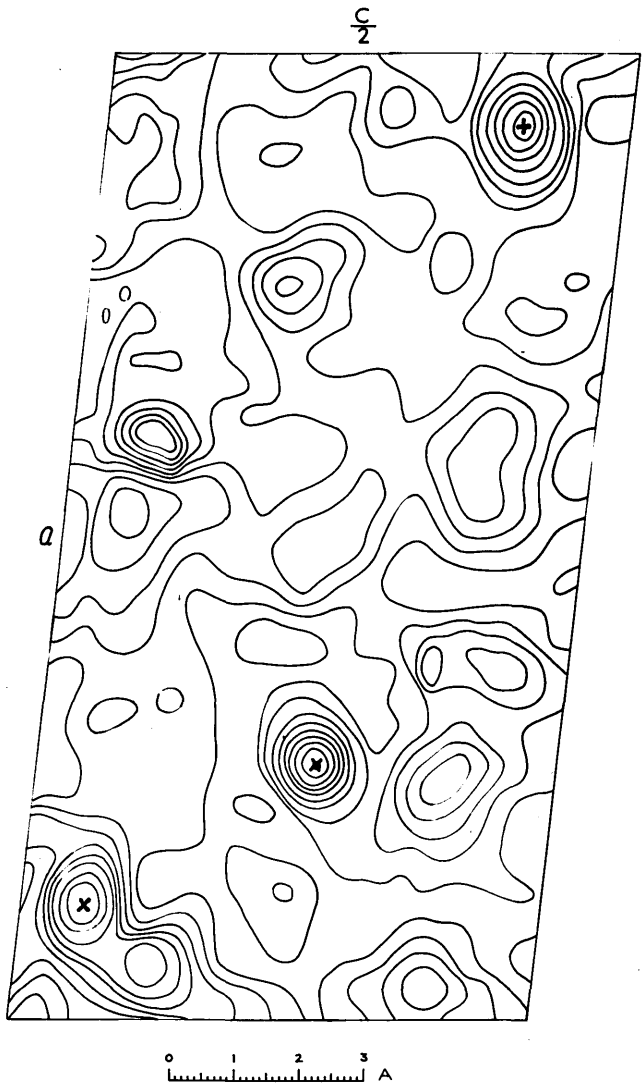


Fig. 5 Harker Section at $y = \frac{1}{2}$.

✗ denotes a Harker peak due to the vector between two symmetry-related iodine atoms.

+ denotes a non-Harker peak due to the vector between two independent iodine atoms.

The equivalent positions for the space group $P2_1$ are (x,y,z) and $(\bar{x}, \frac{1}{2}+y, \bar{z})$. For every pair of atoms related by the screw axis there will be a maximum in the Patterson function at $(2x, \frac{1}{2}, 2z)$. Thus for the space group $P2_1$, the corresponding Harker section is $y = \frac{1}{2}$. $P(x,\frac{1}{2},z)$ may be calculated by means of the expression

$$P(x,\tfrac{1}{2},z) = \sum_h \sum_l C_{hl} \cos 2\pi(hx + lz) \quad \text{.....(23)}$$

where $$C_{hl} = \sum_k (-1)^k |F_{hkl}|^2.$$

The $(h0l)$-$(h7l)$ data provided 2146 $|F_{hkl}|$ values which were used to compute $P(x,\frac{1}{2},z)$, the calculation being performed on Robertson's Universal Fourier Synthesiser (Robertson, 1954, 1955). The resulting Harker section of the Patterson function is shown in Fig. 5. The two peaks corresponding to vectors between the symmetry-related iodine atoms are clearly defined and it should be noted that there is also present a non-Harker peak of appreciable height corresponding to the vector between the two independent iodine atoms. The peak maxima were accurately determined by interpolation along the two axial directions and it was found that the iodine coordinates differed only very slightly from those obtained by a study of the (100) and (010) projections. The coordinates deduced from the Harker section take the following values

Atom	x	y	z
I_1	$\overline{0{\cdot}062}$	$\overline{0{\cdot}037}$	0·029
I_2	0·133	0·037	0·366

6.2. Abortive Fourier Synthesis

With a knowledge of the positions of the iodine atoms it was decided to attempt to solve the structure by successive Fourier syntheses, with the phases calculated only for the iodine contributions as a first approximation. Owing to the very large amount of computation involved the calculations were performed on the Ferranti 'Mercury' computer at Manchester University.

Structure amplitudes and phase angles based on the iodine coordinates were calculated and the discrepancy between observed and calculated structure amplitudes was found to be about 34%. This value is rather lower than one might have expected but it was decided to use these phases in the computation of a three-dimensional Fourier synthesis. The resulting electron-density distribution was plotted out and proved very disappointing. So many peaks were present that it was very difficult to decide which peaks corresponded to genuine atoms and which were spurious.

An attempt, however, was made to correlate this three-dimensional electron-density distribution with its projection on

(010), shown in Fig. 4. The sections of the 3-D electron-density distribution were placed in turn on top of the (010) electron-density projection and peaks common to both selected subject to the following conditions:-

a) if a peak in the 3-D map corresponded to a ridge or peak of electron-density in the (010) projection it was accepted as genuine

b) if two or more maxima in the 3-D map possessed very similar x- and z- coordinates and hence overlapped in the projection they were not accepted for the present purpose.

In total some forty peaks were selected and, assuming them to be carbon atoms, their x- and z- coordinates used to calculate structure factors for the (h0ℓ) data. For each structure factor calculation, the selected peaks were divided into groups and various combinations examined. It was found, however, that the R-factor showed relatively little sign of improvement on testing a series of possible combinations of the selected peaks. The failure of this Fourier synthesis prompted us to consider making use of the Patterson minimum function, at least until some solution of the electron-density problem had been achieved.

6.3. Three-Dimensional Patterson Function

As a preliminary to the study of the minimum function a three-dimensional Patterson synthesis was computed using a programme devised by Rollett for English Electric 'DEUCE'. Since the

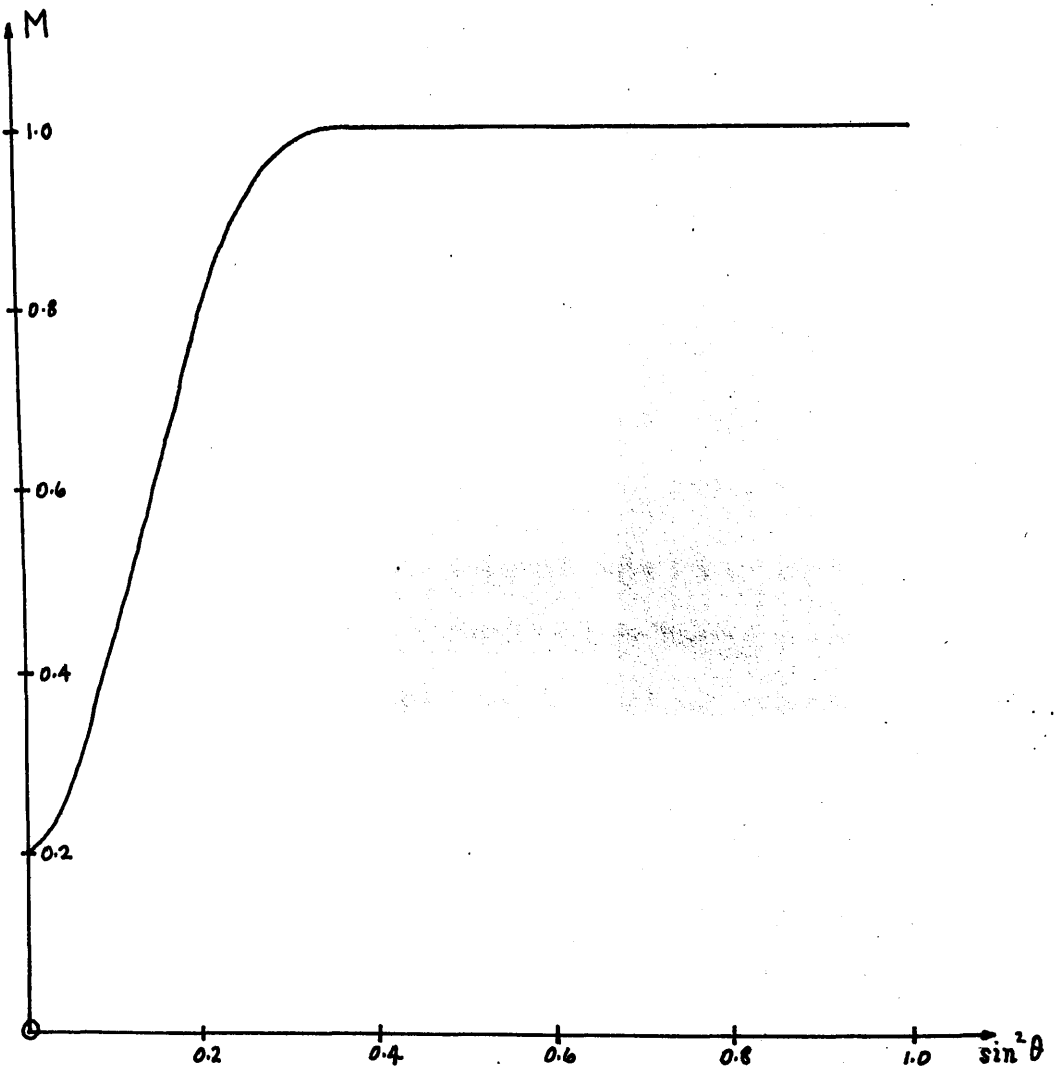


Fig. 6 Modification function used in three-dimensional Patterson synthesis.

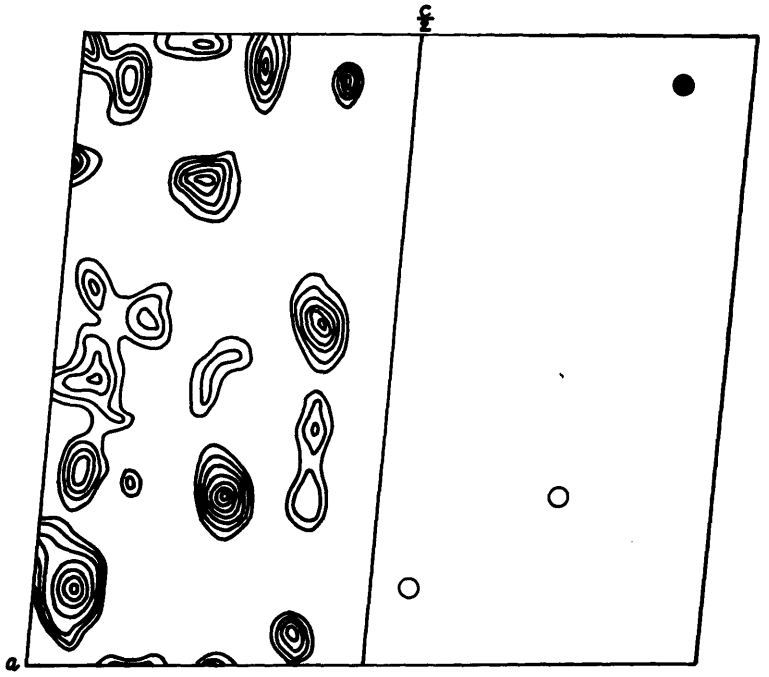


Fig. 7 Harker section of three-dimensional Patterson function.

- O denotes a Harker peak corresponding to the vector between two symmetry-related iodine atoms.
- ● denotes a non-Harker peak corresponding to the vector between two independent iodine atoms.

Patterson function involves quantities proportional to f^2 rather than to f the peaks tend to be broader and less well-defined than the peaks of an electron-density map. Various methods of 'sharpening' the Patterson function have been developed (Lipson and Cochran, 1953) and the general principle involved is to modify the Fourier coefficients, $|F_{hk\ell}|^2$ so as to give more weight to the middle and higher ranges of sin θ and less weight to the low-order terms, the process being roughly equivalent to the use of a smaller temperature factor. The modification function, M, used in the present work is shown in Fig. 6. For the minimisation of computing time the data were arranged so that the number of h values was lowest and the number of k values next. The three-dimensional 'sharpened' Patterson synthesis was computed in sections perpendicular to the b-axis at intervals of $^1/48$ from y=0 to y=$^1/2$. The interval of division along the a- and c-axes was also $^1/48$, the Patterson function being calculated for the intervals x=0 to x=1 and z=0 to z=$^1/2$. The value of the Patterson function at each point was recorded on a tracing paper grid but the contours were not drawn since the Patterson distribution was computed with the main purpose of deducing the values of the minimum function. The coordinates of the iodine atoms were confirmed and it is of interest to inspect the Harker section at y=$^1/2$, shown in Fig. 7. The Harker section illustrated in Fig. 5 was computed using 2146 $|F(hk\ell)|^2$ values, whereas the

section at $y=^{1}/_{2}$ in the 3-D Patterson synthesis was calculated with 2927 $|F(hk\ell)|^2$ values. Comparison of the two sections shows that the inclusion of more data has resulted in a much smaller non-Harker peak relative to the true Harker maxima.

6.4. The Patterson Minimum Function

Many attempts have been made to derive relationships between fundamental sets and vector sets (Wrinch, 1939; Buerger, 1950, 1951; Beevers and Robertson, 1950; Clastre and Gay, 1950 a,b; Garrido 1950 a,b; McLachlan, 1951). These different approaches may be termed "superposition methods" and the chief practical difficulty lies in the fact that the theory is strictly valid only for a vector set of points and not for the Patterson function. However, superposition of Patterson functions leading to composite functions such as addition, product and minimum functions has been of considerable use in the solution of crystal structures.

In the present work the method employed may be summarised thus:-

Let (x, y, z) be the coordinates of 'light' atoms and (x_I, y_I, z_I) the coordinates of the iodine atoms. Then vectors between 'light' atoms and I_1 will be represented by peaks in Patterson space with coordinates

$$\left.\begin{aligned} X &= x - x_{I_1} \\ Y &= y - y_{I_1} \\ Z &= z - z_{I_1} \end{aligned}\right\} \qquad \text{.....(24)}$$

A grid was drawn out representing the (x, z) plane and on it were marked the positions of I_1 and I_2.

$$\text{Now} \quad y_{I_1} = \frac{-8{\cdot}5}{240} \quad \text{and} \quad y_{I_2} = \frac{8{\cdot}5}{240}$$

However, since the Patterson function had been computed in sections at intervals of $y = {}^{5}/_{240}$, it was convenient to assume that

$$y_{I_1} = {}^{-10}/_{240} \quad , \quad y_{I_2} = {}^{10}/_{240}$$

Applying the relationships (24) to the two iodine atoms I_1 and I_2,

$$\left.\begin{aligned} y &= Y + y_{I_1} \\ y &= Y + y_{I_2} \end{aligned}\right\} \quad \text{i.e.} \quad \left.\begin{aligned} y &= Y - {}^{10}/_{240} \\ y &= Y + {}^{10}/_{240} \end{aligned}\right\}$$

Thus, in preparing the superposition function for, say, $y = {}^{15}/_{240}$, the Patterson section $Y = {}^{25}/_{240}$ was placed with its origin at I_1 and the section $Y = {}^{5}/_{240}$ placed with its origin at I_2. The superposition function was plotted for each point of the grid, taking the minimum value of the two Patterson functions at each grid point. In this way the minimum function, M_1, was prepared in sections at $y = {}^{0}/_{240}$ to $y = {}^{120}/_{240}$ at intervals of $y = {}^{5}/_{240}$. In like manner the minimum function, M_2, with respect to iodine atoms

I_1^* and I_2^* was prepared and finally M_1 and M_2 were directly superposed to give the complete minimum function, M. There was then a fairly high probability that peaks in the distribution, M, corresponded to genuine atomic sites of the 'light' atoms and it was hoped to confirm this result when the minimum function, M, was compared with the first three-dimensional electron-density distribution.

6.5. Three-Dimensional Fourier Analysis

Some five to six months after first receiving the 3-D Fourier results from Manchester, and just before the completion of the minimum function, we were informed that there had been a subtle error in the Fourier programme, invalidating the previous results and it was accordingly decided to repeat the Fourier calculation using the corrected version of the programme. It is convenient at this stage to discuss the theory of the weighting system applied to the Fourier coefficients.

Luzzati (1953) has shown that if, with a compound containing heavy atoms, only some of the atoms are used in computing the phase angles then the resolution is poorer in the non-centrosymmetrical case than in the centrosymmetrical case. Woolfson (1956) has recently demonstrated how the resolution may be improved for centrosymmetrical structures by computing a Fourier series with coefficients modified according to the probability that F and F_H have the same sign, where F_H is the structure factor calculated

on the basis of only some of the atoms in the unit cell.

For the non-centrosymmetrical case a method has been developed by Sim (1959) which leads to improved resolution and which may be summarised thus:-

Let $|F_H|$ and α_H be the values of the structure amplitude and phase angle calculated on the basis of only some of the atoms in the unit cell.

The probability of $(\alpha-\alpha_H)$ lying between ξ and $\xi + d\xi$ for a structure factor with fixed values of $|F|$, $|F_H|$ and α_H is given by

$$p(\xi)\, d\xi = \exp(X \cos\xi)\, d\xi / 2\pi I_0(X) \quad \ldots\ldots\ldots\ldots\ldots\ldots(25)$$

where I_0 is the modified zero-order Bessel function (Watson, 1922) and

$$X = 2|F|\,|F_H| \Big/ \sum_{i=1}^{n} f_i^2 \quad ,$$

the summation being over the atoms not used in the evaluation of α_H. The probability that $(\alpha-\alpha_H)$ lies between the limits $\pm\xi$ is then given by

$$P(\xi) = \int_{-\xi}^{+\xi} p(\xi)\, d\xi \quad \ldots\ldots\ldots\ldots\ldots\ldots(26)$$

Values of $P(\xi)$ for various values of X and ξ have been tabulated and it appears that the larger the value of X the more likely $|\alpha - \alpha_H|$ is to be small and vice-versa.

When it is probable that $|\alpha - \alpha_H|$ is small the weighting factor should be large, approaching unity, and when $|\alpha - \alpha_H|$ is likely to be large then the weighting factor should be small. The weighting factor should be a function of $P(\xi)$ and Sim has suggested the use of

$$W = 2P(90^\circ) - 1 \qquad \text{.....(27)}$$

Values of W as a function of X have been calculated and it has been shown that the resolution is improved, especially in the case where the number of atoms used in calculating α_H is small compared to the total number of atoms in the unit cell.

The (hkℓ) data were divided into groups with 2 sin θ values lying in ranges of 0·05 unit and the weighting factor corresponding to the mid-point of the interval evaluated. The observed structure amplitudes weighted according to the above method and the phase angles calculated for the iodine atoms were used to recalculate the three-dimensional electron-density distribution. The Fourier synthesis, which shall be referred to as F1, was computed in sections perpendicular to the b-axis at intervals of $^1/48$ from y = 0 to y = $^1/2$. The interval of division along the a- and c- axes was also $^1/48$ covering the ranges x = 0 to x = 1 and z = 0 to z = 1. The electron-density function was thus evaluated at 55,296 points in the unit cell. These electron-density values were transferred to sheets of tracing paper and the contours drawn. The approximate time

taken by two workers to prepare any one section was from 1½ to 2 hours. The contoured sections were redrawn on sheets of perspex which were then stacked in a metal frame, with perspex spacers, cut to scale to approximately the correct width, placed between the sections. The whole assembly was illuminated from below by strips of fluorescent lighting which greatly facilitated the examination of the electron-density distribution. Careful scrutiny of the electron-density distribution showed that this recalculated Fourier synthesis offered a much better chance of success than the original, the peaks having been reduced to a more manageable number.

Using a different colour of crayon the minimum function was then contoured on the perspex sheets, section for section. The complete electron-density distribution was thoroughly compared with the minimum function and strong peaks which were well-represented in both functions were considered to represent genuine atoms. For the two molecules in the asymmetric unit it proved possible to assign coordinates to 50 atoms other than the two iodine atoms. At this stage the relative peak heights did not justify distinction between carbon and oxygen atoms.

From this point all computational work was done in Glasgow using the programmes developed by Rollett for English Electric 'DEUCE'. Using the carbon scattering factor curve of MacGillavry et al (1955) and the iodine scattering factor curve of James and Brindley (1931), phase angles were calculated on the

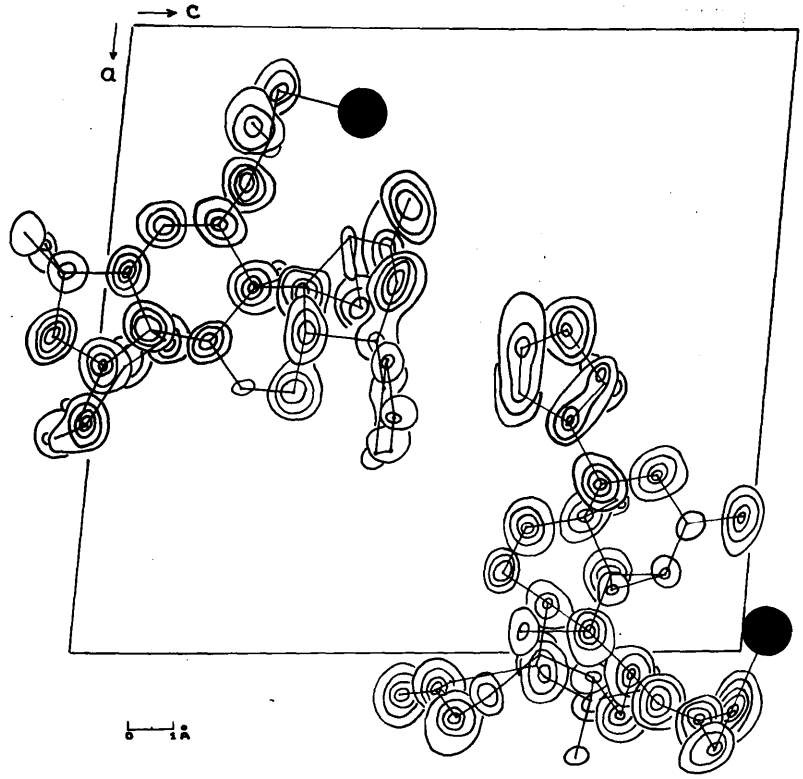


Fig. 8 Part of the electron-density distribution, F2, represented by superimposed contours and showing the two molecules in the asymmetric unit.

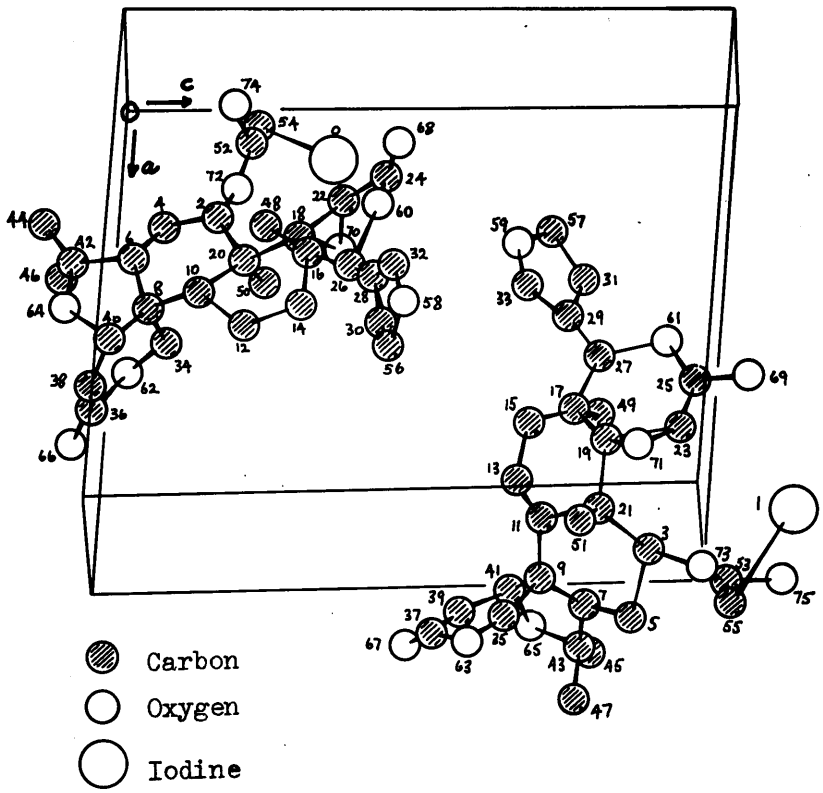


Fig. 9 The arrangement of atoms in the crystal asymmetric unit.

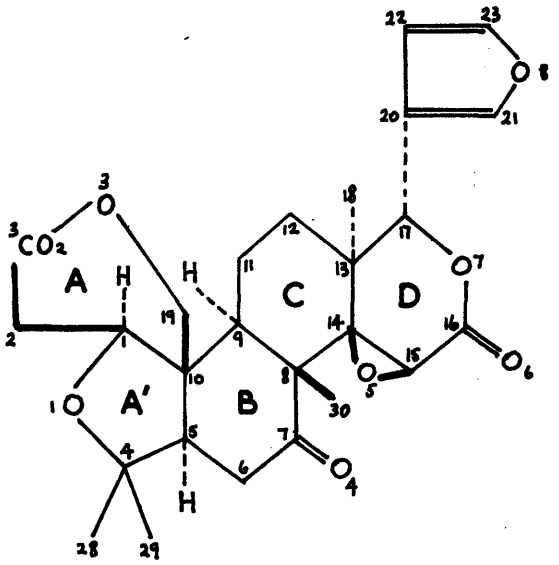


Fig. 10 Molecular structure of limonin.

basis of the fifty 'carbon' atoms and two iodine atoms. The temperature factor employed was $B = 3{\cdot}0$ Å^2 for all 52 atoms. With these phases a second Fourier synthesis, F2, was computed and the sections drawn out on a separate stack of glass sheets. Comparison of F1 and F2 showed that much of the background detail present in F1 had been eliminated in F2. A careful survey of F2 enabled us to specify two, distinct, well-separated molecules, mutually consistent in the chemical sense. Although the presence of two molecules in the asymmetric unit greatly complicated the calculations in this work, the fact that the peak positions were found to conform precisely to two chemically identical but differently oriented molecules provided convincing evidence of the validity of our results.

Part of the electron-density distribution, F2, is shown in Fig. 8, by means of superimposed contours. The arrangement of the atoms in the crystal asymmetric unit is illustrated in Fig. 9. The structure proposed for limonin on the basis of F2 is shown in Fig. 10, the lettering and numbering having been proposed by Professor Barton in accordance with the theories as to the biogenesis of limonin.

On the basis of peak heights it was not possible to distinguish beyond all doubt between oxygen and carbon atoms. In particular, in the lactone ring A it was not possible to rule out the possibility that O(3) might be carbon and C(2) oxygen.

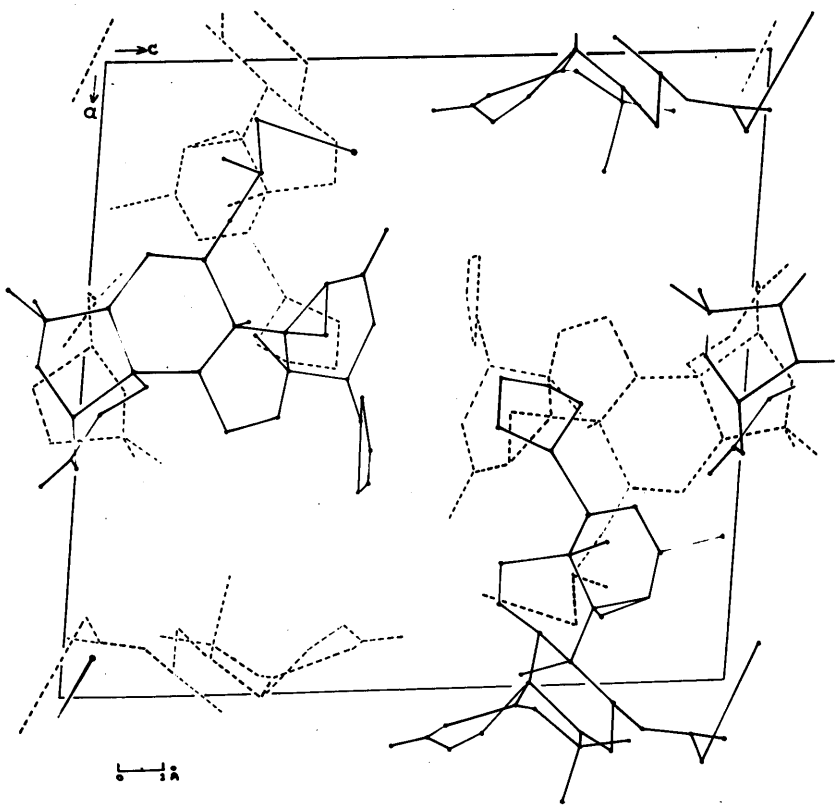


Fig. 11 Atomic positions, as determined by F2, projected on (010) for comparison with Fig. 4.

However, as will be seen later, this ambiguity is eliminated by the chemical evidence.

At this stage the electron-density projection on (010), shown in Fig. 4, was compared with a sketch, drawn to the same scale, of the atomic positions projected on (010). This sketch is illustrated in Fig. 11 and it is of interest to note that about 30 peaks in the electron-density projection correspond to genuine atoms.

7. Refinement of Structure

Refinement of the crystal structure of epi-limonol iodoacetate has been conducted by successive structure factor calculations and three-dimensional Fourier syntheses. After each structure factor calculation the discrepancies, Δ, equal to $(|F_o| - |F_c|)$ were examined and those with Δ positive and greater than or equal to $|F_c|$ were omitted from the corresponding Fourier synthesis. This criterion of selection has resulted in about 2800 structure factors out of 2927 being used for each Fourier synthesis.

Two Fourier syntheses, F3 and F4, were computed and one cycle of least squares analysis performed, omitting the furan ring and C(17) of one molecule from the phasing. The reason for this procedure was that these atoms, although well-resolved, were still fairly weak.

The succeeding Fourier synthesis, F5, used phases based on

Table 2

Atomic coordinates derived from F5

Molecule I

Atom	x/a	y/b	z/c	Atom	x/a	y/b	z/c
0	·135	·035	·366	38	·634	·099	·003
2	·321	·160	·164	40	·543	·143	·010
4	·314	·105	·081	42	·408	·104	·938
6	·395	·130	·034	44	·352	·188	·880
8	·488	·104	·077	46	·628	·485	·085
10	·489	·165	·164	48	·445	·394	·252
12	·581	·181	·220	50	·414	·029	·252
14	·581	·270	·281	52	·173	·161	·216
16	·495	·306	·304	54	·101	·105	·240
18	·429	·217	·297	56	·456	·013	·579
20	·413	·137	·215	58	·312	·034	·561
22	·351	·215	·361	60	·421	·373	·427
24	·350	·318	·410	62	·453	·420	·978
26	·511	·348	·397	64	·497	·121	·931
28	·570	·435	·417	66	·339	·423	·056
30	·677	·434	·417	68	·276	·335	·435
32	·376	·099	·557	70	·438	·145	·379
34	·490	·477	·915	72	·250	·115	·207
36	·376	·481	·011	74	·165	·266	·194

Table 2

Atomic coordinates derived from F5.

Molecule II

Atom	x/a	y/b	z/c	Atom	x/a	y/b	z/c
1	·062	·462	·972	39	·039	·333	·432
3	·027	·127	·835	41	·031	·364	·327
5	·101	·037	·828	43	·909	·333	·223
7	·954	·441	·217	45	·812	·339	·241
9	·999	·474	·306	47	·915	·273	·142
11	·927	·052	·705	49	·747	·089	·788
13	·860	·080	·630	51	·008	·245	·706
15	·795	·166	·650	53	·099	·234	·950
17	·785	·174	·746	55	·145	·333	·971
19	·880	·214	·788	57	·518	·301	·622
21	·957	·164	·754	59	·472	·252	·697
23	·870	·263	·873	61	·714	·303	·837
25	·786	·282	·896	63	·893	·437	·417
27	·722	·281	·757	65	·960	·287	·293
29	·624	·265	·710	67	·911	·288	·500
31	·594	·293	·629	69	·769	·297	·976
33	·553	·236	·753	71	·887	·332	·799
35	·070	·028	·623	73	·070	·226	·871
37	·941	·358	·452	75	·087	·150	·003

Note:- Numbering system as in Fig. 9.

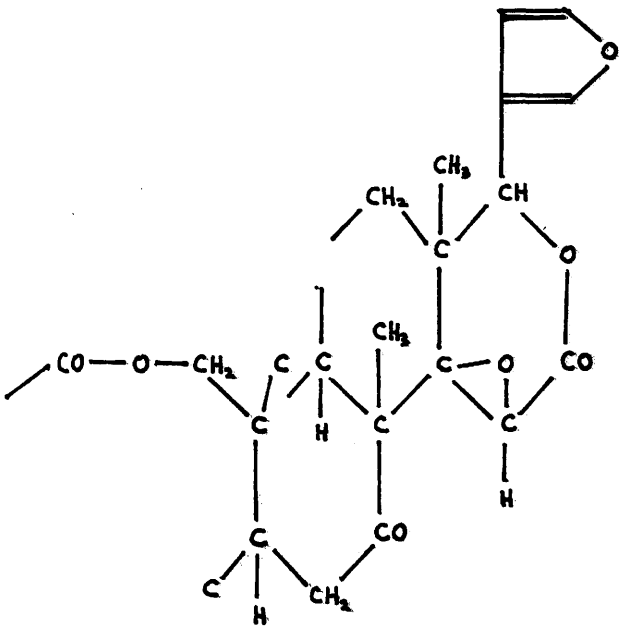
all 76 atoms in the asymmetric unit. The resulting electron-density distribution showed considerable improvement in the furan ring but the oxygen peak was still rather unsatisfactory. The peak heights throughout the whole electron-density distribution were still too high due to an error in the scale factor but the ratio of the peak heights of carbon and oxygen atoms was correct. The scale factor and temperature factor have been adjusted and at the time of writing a sixth Fourier synthesis is being computed. It is hoped that this computation will substantially improve the oxygen peak of the furan ring.

The atomic coordinates obtained from F5 are listed, for the two independent molecules of the asymmetric unit, in Table 2. The structure factors with associated phase angles calculated on the basis of these coordinates are compared with the observed values in Appendix 1. The discrepancy between these observed and calculated structure factors is 19·8%.

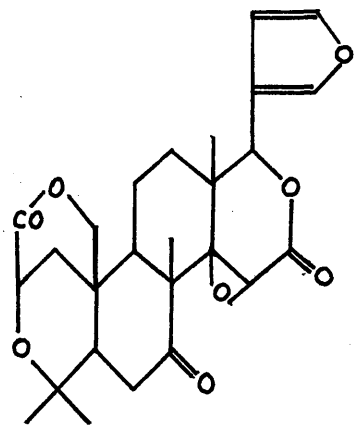
8. Discussion of the Structure

Over the past few years three groups of organic chemists, in London, Zürich and Harvard, have been studying the constitution of limonin and recently the results of their work have been made available to us. These results together with our findings have been submitted for publication in Experientia.

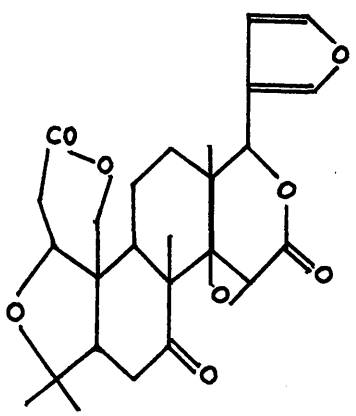
At the time of our solution of the structure the mass of chemical evidence provided conclusive proof for the presence of



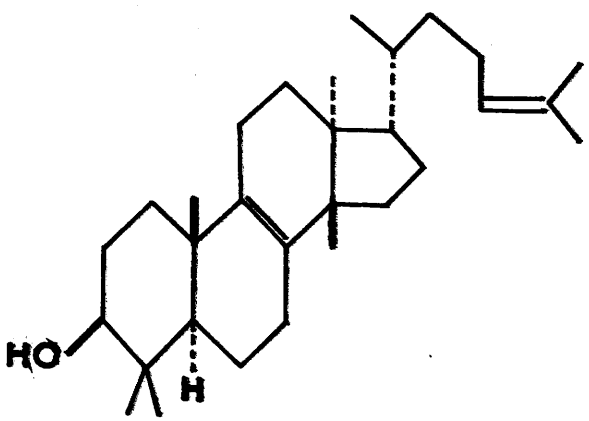
II



III



IV


HO
H

the system II which accounts for 20 of the 26 carbon atoms and 7 of the 8 oxygen atoms. During recent months two possible complete structures III and IV had been considered. More weight had been attached to the latter which corresponds exactly to the structure determined by X-ray methods. No information was available, however, from the chemical evidence, concerning the stereochemistry of limonin.

Biogenetically limonin may be considered to be a tetracyclic triterpenoid of the euphol type, V, (Barton et al, 1954, 1955; Arigoni et al, 1954). According to Barton (private communication) there is a distinct possibility that the structures of several bitter principles may be similar to that of limonin, especially nomilin, $C_{28}H_{34}O_9$, (Emerson, 1948, 1951) and obacunone, $C_{26}H_{30}O_7$, (Sondheimer et al, 1959; Dean and Geissman, 1958).

The stereochemistry of limonin is as illustrated in Fig. 10. In epi-limonol, the hydroxyl group O(4) at C(7) is equatorial. This is in agreement with the experimental observation (Melera et al, 1957; Fujita and Hirose, 1954, 1956) that limonin yields epi-limonol on reduction with sodium amalgam, a process which is known to yield the more stable of a pair of stereoisomers. Thus limonol itself must be the axial isomer in agreement (Barton, 1953) with its mode of formation. The methyl group, C(23), at ring junction B/C is axial. In the trans-decalin system ring B has the chair and ring C the boat conformation.

In so far as the X-ray analysis is concerned the refinement process should terminate fairly soon. It is intended to compute an F_c synthesis, obtain the back-corrections and finally compute a three-dimensional F_o synthesis.

CHAPTER III

THE GENERAL AND MOLECULAR STRUCTURE

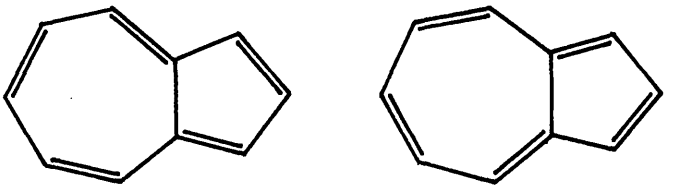
OF AZULENE

CHAPTER III

THE CRYSTAL AND MOLECULAR STRUCTURE OF AZULENE

1. Introduction

Azulene, $C_{10}H_8$, belongs to a group of compounds responsible for the intense blue colour in certain essential oils. The structure of the parent compound was elucidated by Pfau and Plattner (1936) and has been generally represented by two Kekulé-type formulae:-



However, the above structures do not fully explain the properties of azulene. For instance, azulene is basic, dissolving in strong, aqueous acids and it has a dipole moment of 1·0 Debye unit. X-ray studies of this hydrocarbon were initiated with the following aims in view:-

a) an accurate determination of the molecular geometry.

b) a detailed picture of the electron-density distribution with the possibility of detecting any charge-transfer effects.

c) correlation of the results with those obtained from the numerous theoretical studies of azulene.

2. Previous Examinations of Azulene

The earliest study of azulene was made by Misch and van der Wyk (1937) who determined the unit cell parameters and

suggested that the space group was either $P2_1(C_2^2)$ or $P2_1/m(C_{2h}^2)$. Later work by Günthard et al. (1948) revealed the absence of the (h0ℓ) spectra for h = 2n+1 in addition to the (0k0) absences for k = 2n+1, and led these workers to propose $P2_1/a(C_{2h}^5)$ as the space group. With two molecules in the unit cell this would require that the molecule possesses a centre of symmetry, which is in conflict with the chemical evidence. The explanation put forward was that the centre of symmetry might be a statistical effect due to the random reversal of direction of the moleculesin the crystal lattice. Later Günthard (1949) suggested that the (h0ℓ) and (0k0) halvings might be attributed to pseudo-symmetry and that the true space group was in fact $P2_1(C_2^1)$, $Pa(C_s^2)$ or even $P\bar{1}(C_i^1)$.

More recently Takeuchi and Pepinsky (1956) reported the results of their preliminary analysis of azulene. They considered the two possible cases:-

a) the random arrangement in $P2_1/a$ and

b) the space group Pa with the projection of molecules on the b-axis symmetric about a point lying mid-way between the glide planes and thus excluding (0k0) spectra when $\underline{k}$ is odd.

Successive structure factor calculations and Fourier projection reduced the R-factor for the (h0ℓ) data to less than 20% in the space group Pa, but failed to reduce it below 24% in $P2_1/a$.

Table 3.

Unit cell parameters.

a	$7{\cdot}884 \pm 0{\cdot}008$ Å
b	$5{\cdot}988 \pm 0{\cdot}008$ Å
c	$7{\cdot}840 \pm 0{\cdot}008$ Å
β	$101^{\circ}33' \pm 20'$
V	$362{\cdot}6$ Å^3
$\rho_{obs.}$	$1{\cdot}175$

Number of molecules per unit cell = 2

$\rho_{calc.}$ $1{\cdot}174$

Absorption coefficient for CuKα X-radiation ($\lambda = 1{\cdot}542$ Å) = $6{\cdot}06$ cm^{-1}

Absent Spectra:- (h0ℓ) when h = 2n+1

(0k0) when k = 2n+1

2.1 Two-Dimensional Studies

For the sake of continuity a short account is now given of the two-dimensional refinement and collection of three-dimensional data carried out by Robertson and Shearer (1956).

Unit cell parameters were calculated from zero-layer Weissenberg photographs, standardised with sodium chloride powder lines, about the a-, b-, and c- axes and the values obtained are listed in Table 3.

Determination of the Space Group

The absences quoted in Table 3 would indicate the space group $P2_1/a(C_{2h}^5)$ and would require the random arrangement mentioned earlier. In this space group the projection of the structure on (010) is centrosymmetrical and to test this situation the statistical methods of Wilson (1949) and Howells (1950) were applied to the (h0ℓ) spectra. The results of the N(z) test indicated with apparently a fair degree of certainty that the b-axis projection is non-centrosymmetrical, implying that the (0k0) absences are accidental and that the space group is $Pa(C_s^2)$. This deduction is in agreement with the observation by Bernal (1956) that azulene crystals show {001} and {110}, but in addition each crystal shows one and only one form of the plane {100}, indicating that azulene possesses no axis of symmetry.

Structure Determination

Proceeding on this assumption a trial structure was postulated, based on the chemical formula, placing the two molecules with respect to the glide plane in such a way that an almost exact halving of the spectra was obtained. Refinement of the (010) projection by difference syntheses reduced the R-factor to 11·1%. An electron density projection of the structure on (010) is shown in Fig. 13. It is seen that the molecular plane is steeply inclined to (010) and this prevents perfect resolution of all the atoms. Refinement of the (100) projection by difference syntheses gave an R-factor for the (0kℓ) data of 17·2% and further improvement was obtained by refinement of the (h2ℓ) data using difference generalised projections.

2.2 Recording of Three-Dimensional X-Ray Data

A photographic survey of points lying within the limiting sphere of the reciprocal lattice was made using CuKα radiation. Reflections were obtained by rotation of crystals of fairly uniform cross-section about the [100], [010], [001] and [110] axes, using an instrument of the equi-inclination Weissenberg type. The multiple film technique (Robertson, 1943) was used to correlate strong and weak reflections, the intensities of strong reflections being derived from small crystals rotated about the [100] and [010] axes and previously dipped in liquid air to reduce the effects of

Table 4.

Layer lines recorded and dimensions of crystals used.

Layer lines	Dimensions of crystal specimen (mm.) Cross-section x length along rotation axis
$(0k\ell)$ - $(5k\ell)$	0·66 x 0·69 x 0·82
$(h0\ell)$ - $(h3\ell)$	0·48 x 0·51 x 0·80
$(hk0)$ - $(hk5)$	0·63 x 0·66 x 0·54
$(hh\ell)$ - $(h,h\pm5,\ell)$	0·60 x 0·66 x 1·02
$(0k\ell)$ - $(2k\ell)$	0·16 x 0·12 x 0·51
$(h0\ell)$ - $(h2\ell)$	0·18 x 0·18 x 0·54

extinction. A list of the layer lines recorded, with the dimensions of the crystal specimens employed, is given in Table 4.

3. Three-Dimensional X-Ray Analysis

3.1 Measurement and Correction of Intensities

The intensities were estimated visually by two independent observers (Watson and Shearer), the ratio of the strongest to the weakest intensity in any layer line being about 26,000: 1. The reduction factor from film to film was determined by the method outlined by Rossmann (1956).

The values of the structure amplitudes were derived by the usual formula for a mosaic crystal:-

$$F^2 = I \sin 2\theta/(1 + \cos^2 2\theta) \quad \ldots\ldots\ldots\ldots \quad (28)$$

Lorentz and polarisation factors were applied, together with the rotation factor appropriate to equi-inclination Weissenberg photographs (Tunell, 1939). Absorption corrections to allow for the shapes of the crystal specimens were obtained by calculating the mean path length for the ray through the crystal as proposed by Albrecht (1939), allowance being made in the case of upper layer lines for the increase in path length and change in Bragg angle for the reflections.

Structure amplitudes for planes whose reflections appeared on different layer lines and which were derived from different

crystal specimens were correlated by reflections common to the different sets. The final values adopted were the average of the different sets after correlation, but when variations occurred among very strong reflections the highest value was generally adopted to offset the possibilities of errors due to extinction.

It has been shown by Whittaker and Robinson (1944) that if there are $\underline{n}$ observations of F_i, the standard deviation in any one observation is

$$\sigma(F_i) = \sqrt{\sum_i (\bar{F} - F_i)^2 / (n-1)} \qquad \ldots\ldots\ldots\ldots (29)$$

where $$\bar{F} = \frac{1}{n} \sum_i F_i$$

This relation clearly holds only for large values of $\underline{n}$ and when $\underline{n}$ equals 2, 3, 4, 5, the significance of $\sigma(F_i)$ becomes rather dubious. However, $\sigma(F_i)$ was evaluated for all the $|F|$ values observed more than once and the standard deviation in the structure amplitude was found to be approximately a constant percentage of the structure amplitude, with $\sigma(F_i) \approx 0.07 |F_i|$.

4. Structure Analysis

4.1 Least Squares Analysis

The coordinates obtained from the projections studied by Robertson and Shearer (1956) were submitted to Professor Pepinsky and Dr Vand of Pennsylvania State University for least squares analysis. Isotropic temperature factors of the form $\exp\left\{-B(\sin\theta/\lambda)^2\right\}$ were provided, with $B = 5{\cdot}5\text{Å}^2$ for both carbon and hydrogen atoms. 670 planes were employed in the analysis and refinement was carried out only on the carbon atoms, the hydrogen contributions being included in the structure factor calculations. Three cycles of least squares refinement, using an IBM 704 computer resulted in a fall of only 1% in the R-factor from 23·4% to 22·4%. Consideration of these results indicated that there was something fundamentally wrong with the structure assigned to azulene in spite of the relatively good agreements in the structure factors for the three main crystallographic zones.

4.2 Space Group Considerations

As indicated earlier it had been assumed that the (0k0) halving was accidental and the azulene molecule was placed in such a position with respect to the glide plane that an almost exact halving of these spectra took place. There was no evidence

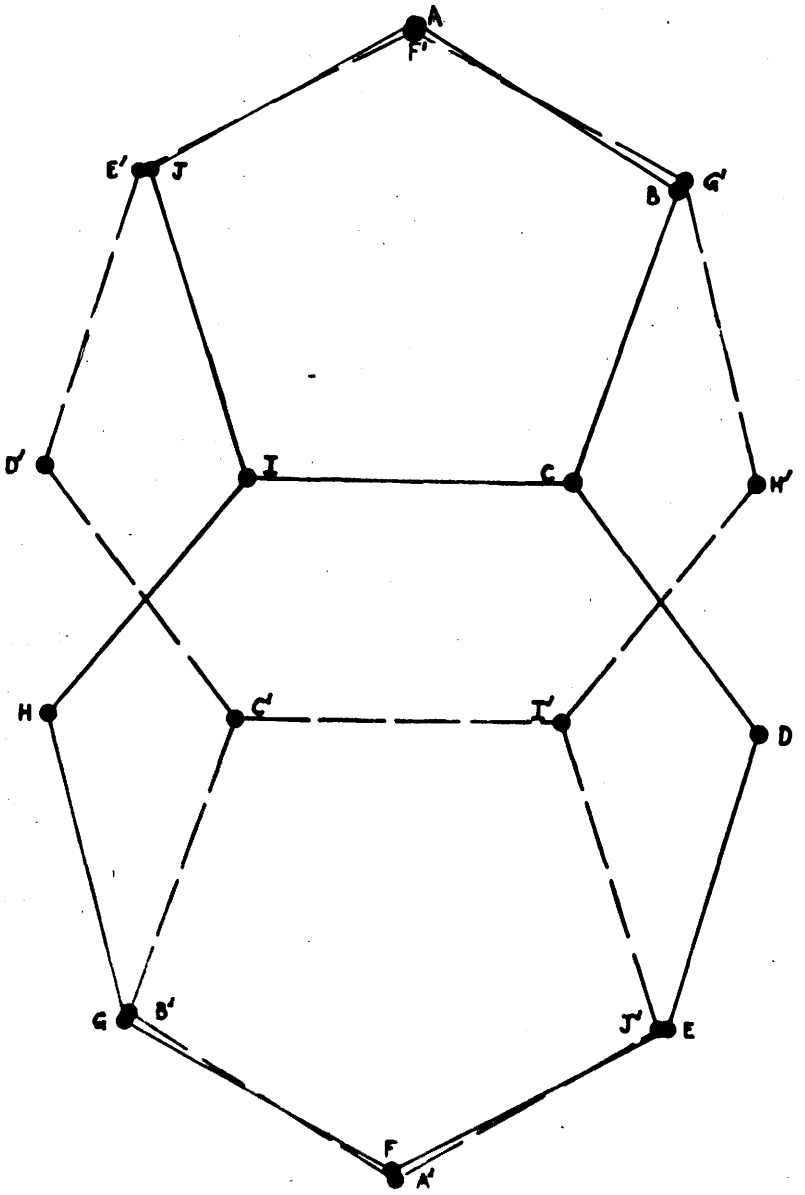


Fig. 12 Superposition of two azulene molecules, the second having been inverted through the symmetry centre.

of disorder in the X-ray photographs but thermodynamical studies by Günthard (1949,1955) have shown that the entropy of solid azulene is substantially higher than that of the isomeric naphthalene, and Günthard has suggested that there might be a statistical disorder effect in crystalline azulene.

The possibility was next considered that the space group might after all be $P2_1/a$, as indicated by the spectral absences and the structure really centrosymmetric. This implies, of course, that the structure is disordered. The atomic coordinates deduced previously were referred to a new origin at the centre of the molecule which was taken to coincide with a centre of symmetry in $P2_1/a$, and the atoms given half-weight. This procedure postulates the superposition of two molecules in the original orientation except for the inversion of the second molecule through the symmetry centre. Physically, and according to the chemical evidence this must correspond to a simple random reversal of direction of successive azulene molecules in the crystal. A view of the resulting superposition of two molecules is shown in Fig. 12.

This disorder is of the same type as found in p-chlorobromobenzene (Hendricks, 1933; Klug, 1947) and is also similar to that reported in the case of 2-amino-4-methyl-6-chloropyrimidine (Clews and Cochran, 1948), though in these cases, because of the symmetry of the molecules, there are effectively only two atoms involved in the disorder. The X-ray photographs of these compounds as of azulene

Table 5.

R-factors for space groups $P2_1/a$ and Pa.

	$P2_1/a$	Pa.
0kℓ	14·9%	20·0%
h0ℓ	10·2%	11·5%
hk0	18·1%	14·2%
hkℓ	20·4%	22·4%

show no trace of additional layer lines and no diffuseness in the spectra, so that the possibility of a fault structure with blocks of the crystal containing molecules properly aligned in one direction and other blocks with molecules properly aligned in the other direction can be excluded, (Robertson et al., 1958).

5. Preliminary Refinement of Centrosymmetrical Structure

Full three-dimensional structure factors were computed for the new centrosymmetric structure at the National Physical Laboratory and the discrepancies are listed in Table 5 for comparison with the non-centrosymmetric values. At this stage the data was sent to Vand and Pepinsky for three-dimensional least-squares refinement and in the meantime some refinement of the projections was attempted.

5.1. Refinement of (010) Projection

This projection belongs to the plane group p2 and the electron density equation is given by

$$\rho(XZ) = \frac{1}{A_c}\Bigg[F(00) + 2\sum_{h=1}^{\infty} F(h0)\cos 2\pi hX + 2\sum_{\ell=1}^{\infty} F(0\ell)\cos 2\pi \ell Z + 2\sum_{h=1}^{\infty}\sum_{\ell=1}^{\infty}\left\{F(h\ell)\cos 2\pi(hX+\ell Z) + F(\bar{h}\ell)\cos 2\pi(-hX+\ell Z)\right\}\Bigg] \quad \ldots\ldots\ldots\ldots (50)$$

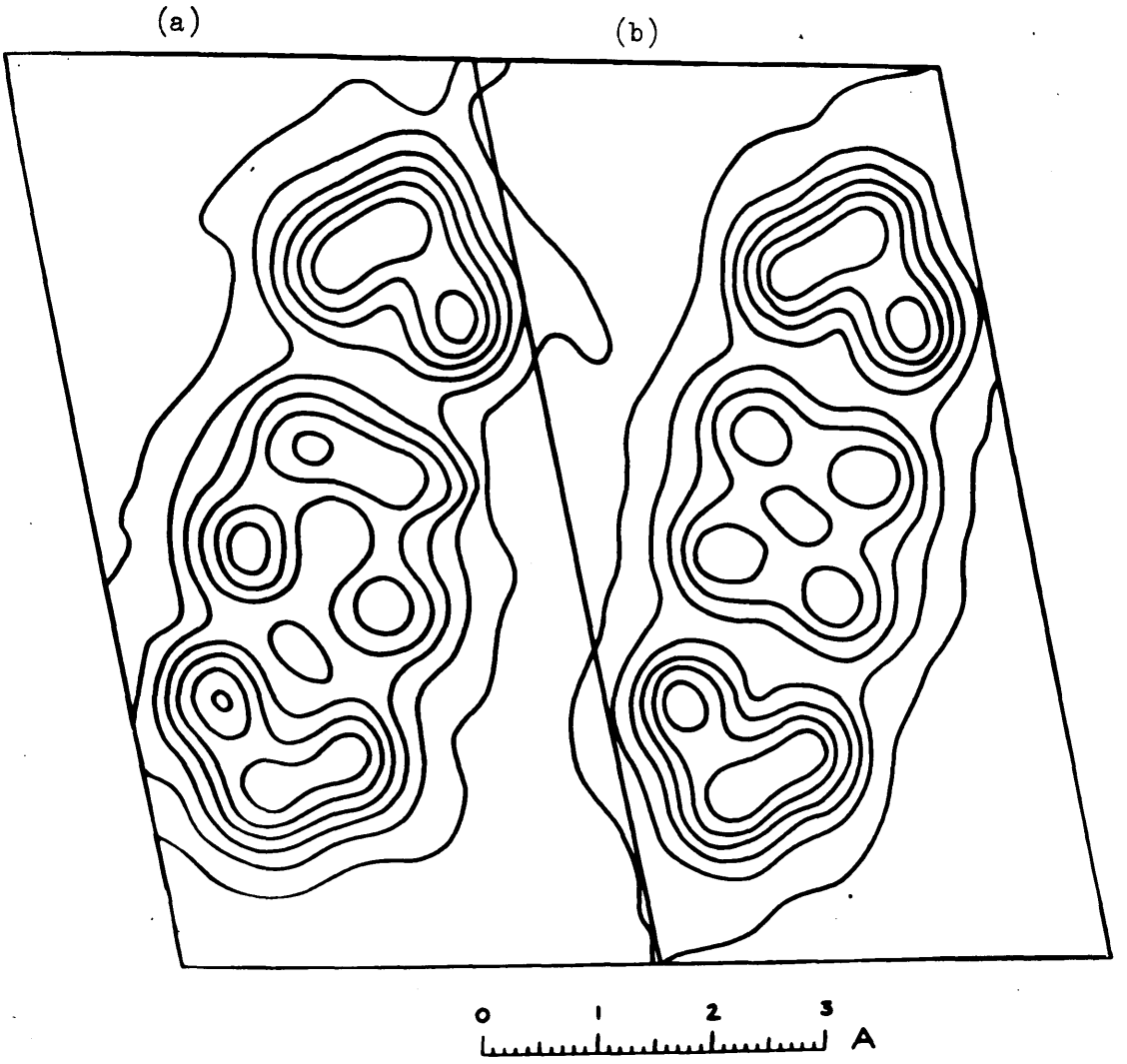


Fig. 13 (a) Projection of electron density on (010) for space group Pa.

(b) Projection of electron density on (010) for space group $P2_1/a$.

Table 6.

Variation of R-factor with k.

Zone	n*	R(Pa)	R($P2_1/a$)
h0ℓ	69	11·5%	10·2%
h1ℓ	124	14·8%	12·1%
h2ℓ	128	22·0%	20·2%
h3ℓ	117	29·1%	26·2%
h4ℓ	88	38·5%	34·7%
h5ℓ	68	36·0%	40·1%
h6ℓ	52	42·0%	45·2%
h7ℓ	25	52·6%	53·2%

* n is the number of reflections, excluding non-observed terms.

78 terms were used in the Fourier summation, with the calculated signs, and the resulting electron density map is shown in Fig. 13 together with the (010) Fourier projection for the space group Pa. The strong similarity between the two electron distributions is very marked and accounts for the good structure factor agreement obtained for the non-centrosymmetric structure.

5.2 Refinement of (100) Projection

The full three-dimensional data were examined to find whether there was any significant variation in the R-factor with $\underline{k}$ and the results are listed in Table 6. It can be seen that as $\underline{k}$ increases so does R, but it must be borne in mind that the number of spectra observed in each zone decreases fairly rapidly with increase in $\underline{k}$. Nevertheless it did seem appropriate to attempt a refinement of the y-coordinates.

The electron density projected along the a-axis is given by

$$\rho(YZ) = \frac{1}{A_c}\left[F(00)+2\left\{\sum_{k=2}^{\infty} F(k0)\cos 2\pi kY + \sum_{\ell=1}^{\infty} F(0\ell)\cos 2\pi \ell Z\right\}\right.$$

$$\left. +4\left\{\sum_{k=2}^{\infty}\sum_{\ell=1}^{\infty} F(k\ell)\cos 2\pi kY \cos 2\pi \ell Z - \sum_{k=1}^{k=2n}\sum_{\ell=1}^{k=2n+1} F(k\ell)\right.\right.$$

$$\left.\left. \sin 2\pi kY \sin 2\pi \ell Z\right\}\right] \quad \ldots\ldots\ldots \quad (31)$$

The Fourier series was summed with 59 terms and the resulting electron

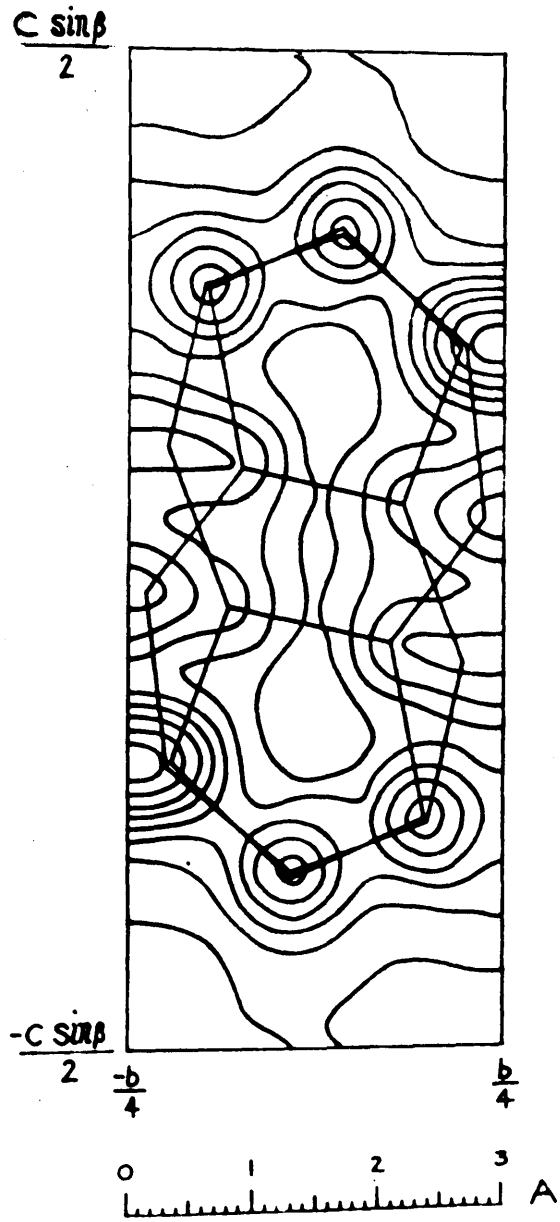


Fig. 14 Projection of electron density on (100).

density map is shown in Fig. 14. Refinement of the (0kℓ) data was then attempted using the least squares method with unit weights applied. Atom pairs A-F', B-G', C-H', I-D' and J-E' (see Fig. 12) were considered to allow for the interaction between atoms and for each pair, two simultaneous linear equations were set up, of the form

$$\left.\begin{aligned} \Delta y_A \sum_u \left(\frac{\partial F_c}{\partial y_A}\right)^2 + \Delta y_{F'} \sum_u \left(\frac{\partial F_c}{\partial y_A}\right)\left(\frac{\partial F_c}{\partial y_{F'}}\right) &= \sum_u (F_o - F_c)\frac{\partial F_c}{\partial y_A} \\ \Delta y_A \sum_u \left(\frac{\partial F_c}{\partial y_A}\right)\left(\frac{\partial F_c}{\partial y_{F'}}\right) + \Delta y_{F'} \sum_u \left(\frac{\partial F_c}{\partial y_{F'}}\right)^2 &= \sum_u (F_o - F_c)\frac{\partial F_c}{\partial y_{F'}} \end{aligned}\right\} \quad \ldots\ldots(32)$$

It was found that, due to the "extreme closeness" of the atoms of each pair, the two equations were almost identical and so gave results without physical meaning. Accordingly simple equations were used, of the form

$$\Delta y_A = \sum_u (F_o - F_c)\frac{\partial F_c}{\partial y_A} \Big/ \sum_u \left(\frac{\partial F_c}{\partial y_A}\right)^2 \qquad \ldots\ldots\ldots\ldots (33)$$

These gave the directions of shift and correspondingly small corrections were made to the y-coordinates resulting in a decrease

in the R-factor for the (0kℓ) data from 14·9% to 13·4% and in the (hk0) data from 18·1% to 16·8%.

An (F_o-F_c) synthesis was then calculated for the (100) projection. Because of the strong overlap of atoms in the electron density distribution it was decided not to estimate the shifts very accurately by the gradient method, but the directions were obtained and approximate correction applied to the y-coordinates. These new coordinates gave R-factors of 11·3% for the (0kℓ) zone and 15·7% for the (hk0) zone.

6. Further Tests of Correct Assignment of Space Group

6.1 Pyroelectric Test

When a crystal belonging to a non-centrosymmetric space group is heated or cooled it develops electric charges, becoming positive at one end and negative at the other. Various qualitative tests have been developed to detect these charges (Wooster, (1949); Robertson, (1935); Martin, (1931); Maurice, (1930).

The method adopted for azulene was to place two small crystals on a piece of aluminium foil which was then suspended in liquid air for three to four minutes. On removal the foil was tapped to determine whether or not the crystals adhered to the foil. Over a series of experiments it could be concluded that azulene crystals did not adhere to the foil, resorcinol and sorbic acid

being used as standards. However, the fact that crystals of azulene showed no detectable pyroelectric effect does not constitute absolute proof that azulene belongs to a centrosymmetrical space group because the pyroelectric effect might be so feeble as to escape detection.

6.2 Theory of Statistical Tests

Wilson Ratio Test

Wilson (1949) has shown that the average value of the intensity $I_{hk\ell}$ of an X-ray reflection, taken for a sufficient range of values of $hk\ell$, is given by

$$\langle I_{hk\ell} \rangle = \sum_{j=1}^{N} f_j^2 \equiv \sum \qquad \ldots\ldots\ldots\ldots (34)$$

where f_j is the atomic scattering factor of the j^{th} atom and the summation is over all the atoms in the unit cell. The probability $P(I)dI$ that $I_{hk\ell}$ should lie between I and I+dI is given, for non-centrosymmetrical and centrosymmetrical space groups respectively, by the relations

$$P_1(I) = \sum^{-1} \exp\left\{ -I/\sum \right\} \qquad \ldots\ldots\ldots\ldots (35)$$

$$P_{\bar{1}}(I) = (2\pi\sum I)^{-\frac{1}{2}} \exp\left\{ -I/2\sum \right\} \ldots\ldots\ldots\ldots (36)$$

The corresponding mean values of the structure amplitudes are

$$\langle |F| \rangle_1 = \sum^{-1} \int_0^{\infty} \sqrt{I} \exp\left\{ -I/\sum \right\} dI = \tfrac{1}{2}(\pi\sum)^{\frac{1}{2}} \quad \ldots(37)$$

$$\langle |F| \rangle_{\bar{1}} = (2\pi\sum)^{-\frac{1}{2}} \int_0^{\infty} \exp\left\{ -I/2\sum \right\} dI = (2\sum/\pi)^{\frac{1}{2}} \quad \ldots(38)$$

Thus $$\rho_1 = \langle |F| \rangle_1^2 / \langle I \rangle \doteq 0{\cdot}785 \quad \ldots\ldots\ldots\ldots(39)$$

$$\rho_{\bar{1}} = \langle |F| \rangle_{\bar{1}}^2 / \langle I \rangle \doteq 0{\cdot}637 \quad \ldots\ldots\ldots\ldots(40)$$

N(z) Distribution

Howells, Phillip and Rogers (1950) have shown that the fraction of all the reflections (other than those systematically absent) of which the intensities are less than or equal to z times the average intensity, $\langle I \rangle$, is, for the non-centrosymmetric case

$$_1N(z) = 1 - \exp(-z) \quad \ldots\ldots\ldots\ldots(41)$$

and for the centrosymmetric case

$$_{\bar{1}}N(z) = 2\pi^{-\frac{1}{2}} \int_0^{(z/2)^{\frac{1}{2}}} \exp(-x^2)dx = \operatorname{erf}(z/2)^{\frac{1}{2}} \quad \ldots\ldots(42)$$

These functions have been tabulated and show that there is a higher proportion of weak reflections for the centrosymmetric case.

N(x) Distribution

Instead of analysing the intensity values a method has been developed by Sim (private communication) in which an entirely analogous procedure is carried out on the structure amplitudes. A possible advantage of this method lies in the fact that an average value of $|F_0|$, applied to all the reflections in a given range of sin θ, will give a closer approximation to the true values of $|F| / \langle |F| \rangle$ than does $\langle I \rangle$ for values of $I/\langle I \rangle$ in the N(z) test because here we are concerned with the variation of f against sin θ as opposed to f^2 with sin θ in the N(z) test. It can be shown that the fraction of all the reflections of which the $|F_0|$ values are less than or equal to x times the average, $\langle |F_0| \rangle$, is, for the non-centrosymmetric case

$$ {}_1N(x) = 1 - \exp(-\pi/4\, x^2) \qquad \ldots\ldots\ldots\ldots\ldots(43) $$

and for the centrosymmetric case

$$ {}_1N(x) = 2\,\varphi\left(\frac{2x}{\sqrt{2\pi}}\right) \qquad \ldots\ldots\ldots\ldots\ldots(44) $$

where $\varphi = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-x^2/2)dx.$

Variance Test

It has been shown by Wilson (1951) that in cases where statistical tests based on distribution functions have failed, it is often

better to use the variance test.

The variance is the mean square deviation of I from its average.

i.e. Variance $= \langle (I - \langle I \rangle)^2 \rangle$ .....(45)

It can be shown that, for a non-centrosymmetrical distribution

$$\langle (I - \langle I \rangle)^2 \rangle = \langle I \rangle^2 - \sum_{i=1}^{N} f_i^4 \quad .....(46)$$

and for a centrosymmetrical distribution

$$\langle (I - \langle I \rangle)^2 \rangle = 2\langle I \rangle^2 - 3 \sum_{i=1}^{N} f_i^4 \quad .....(47)$$

Since $\sum_{i=1}^{N} f_i^4$ is of the order of $\langle I \rangle^2/N$, it is negligible if the unit cell contains an appreciable number of atoms. Thus, for a non-centrosymmetrical distribution,

$$\frac{v_1}{\langle I \rangle^2} = 1 \quad .....(48)$$

and for a centrosymmetrical distribution,

$$\frac{v_{\bar{1}}}{\langle I \rangle^2} = 2 \quad .....(49)$$

6.3 Application of Statistical Tests to Azulene

Wilson Ratio Test

The Wilson ratio, deduced from the (h0ℓ) data, was evaluated to be 0·547 whereas the values corresponding to space groups Pa and

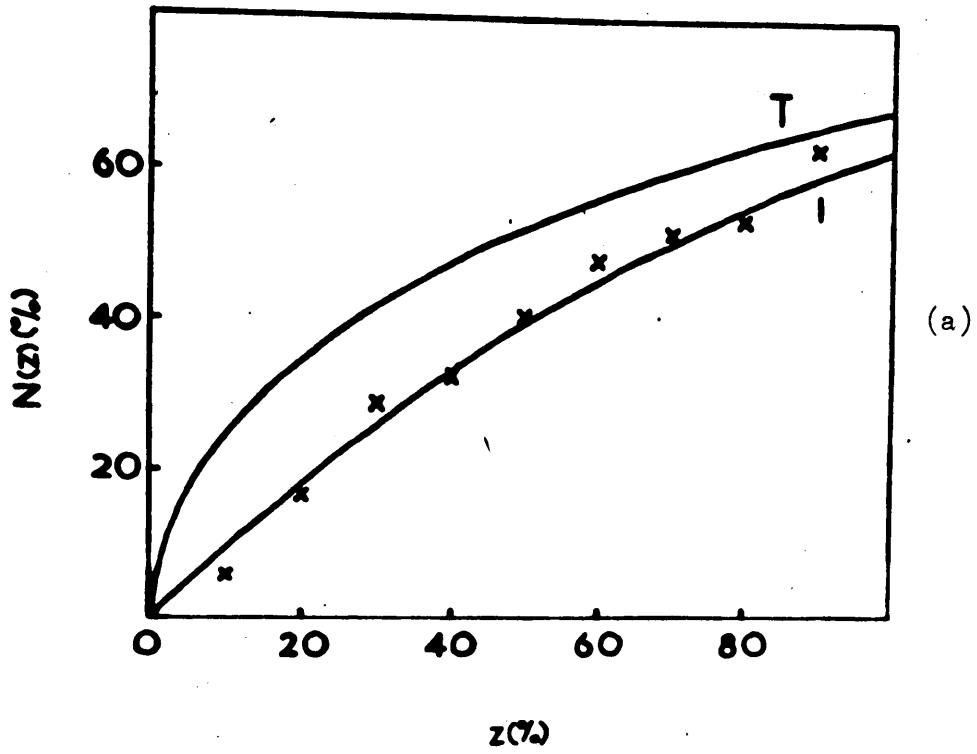


Fig. 15 (a) N(z) test on (h0ℓ) data (results of Shearer and Robertson)

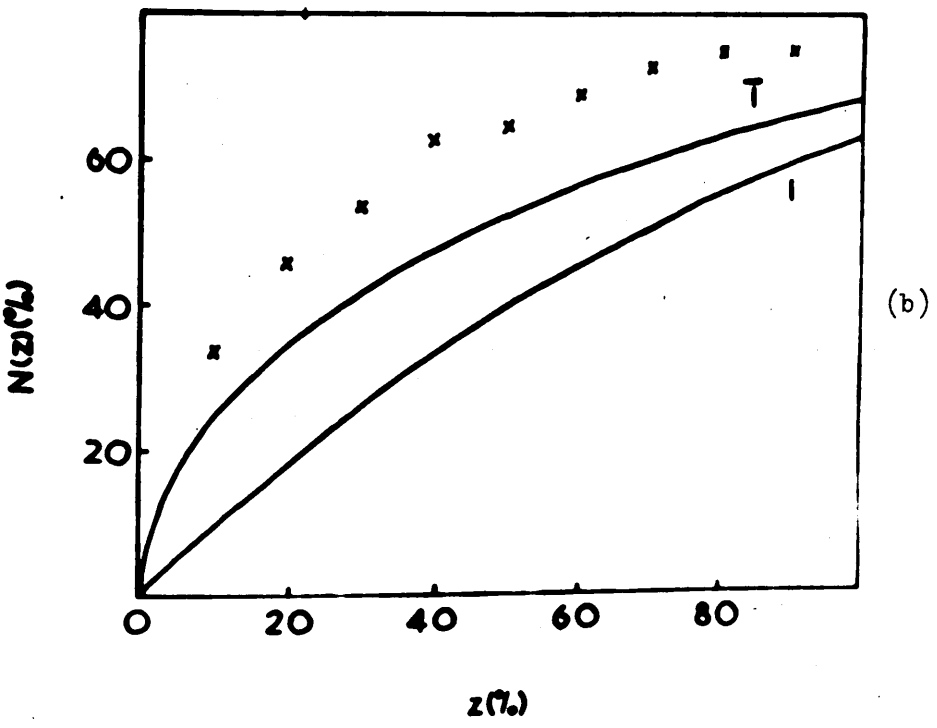


Fig. 15 (b) N(z) test on (h0ℓ) data (results of Watson)

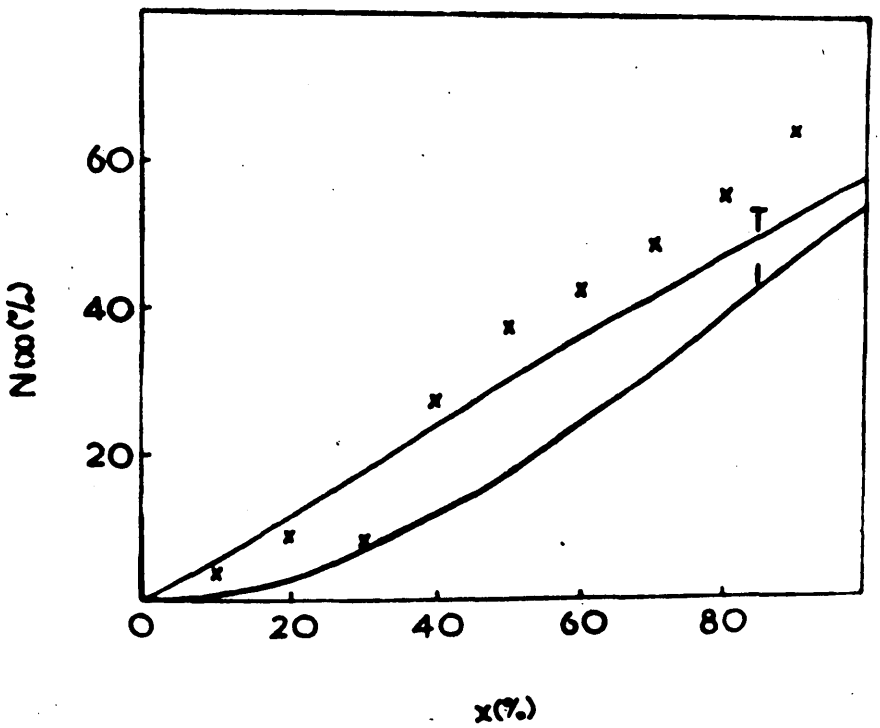


Fig. 16 Results of the N(x) test performed on the (h0ℓ) data.

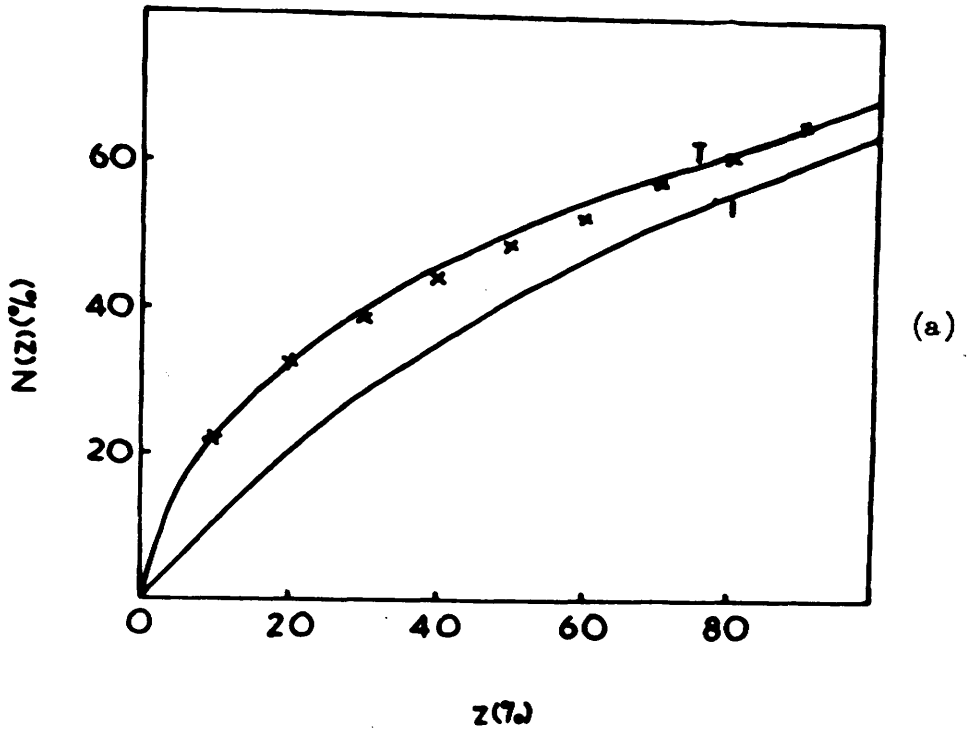


Fig. 17 (a) 3-D N(z) test (method of Sim)

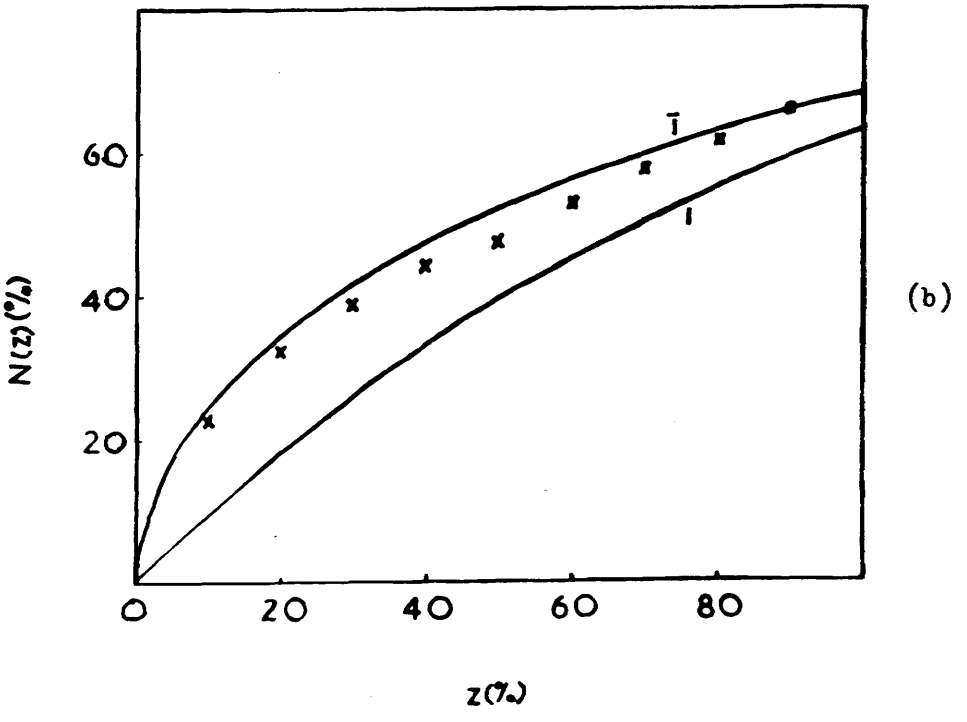


Fig. 17 (b) 3-D N(z) test (method of Howells et al, 1950).

for the same reasons that the N(z) distribution proved inconclusive.

Variance Test

This test was carried out using the three-dimensional data, the spectra being divided into five intervals with a 2 sin θ range of 0·2 unit. The results obtained are listed in Table 7. The experimental specific variance is slightly higher than the theoretical value, 2·0, for a centrosymmetric structure but this can be attributed to the errors in the intensities and to the error in assuming $\sum_{i=1}^{N} f_i^2$ as the mean value of I. However, this test appears to indicate fairly strongly that azulene belongs to a centrosymmetrical space group.

Three-Dimensional N(z) Test

A variation of the N(z) test has been applied to the azulene data by Sim (private communication). Firstly, the N(z) distribution was evaluated in the usual way for some 423 spectra divided into five ranges according to their 2 sin θ values. "Theoretical" N(z) functions were then calculated for the two possible space groups Pa and $P2_1/a$ using the respective F_c^2 values as "intensities". The average standard deviation in a value of N(z) for the calculated distributions is 0·014. The results are shown in Fig. 17.

Combination of the results of this test and the variance test lead us to feel that the assignment of the centrosymmetric space group $P2_1/a$ is correct.

Table 8

Course of refinement of isotropic diagonal least-squares analysis.

Cycle	R	$\sum \Delta^2$
1	91·0%	18921
2	21·0%	2035
3	17·0%	1351
4	15·3%	1148
5	14·3%	1034
6	13·7%	990
7	13·8%	983
8	13·5%	976

Table 9

Parameter shifts indicated by isotropic diagonal least-squares analysis.

	Initial average Shift	Final average Shift	Final maximum Shift
Δx(Å)	0·013	0·003	0·007
Δy(Å)	0·029	0·004	0·012
Δz(Å)	0·011	0·002	0·005
ΔB(Å^2)	0·265	0·091	0·190

7. Three-Dimensional Refinement of Centrosymmetric Structure.

7.1. Isotropic Diagonal Least-Squares Refinement

7.1.1. Computational Details

Least-squares refinement of the disordered structure was performed for us by Professor R. Pepinsky using the IBM 704 programme NY XR1. This programme does not compute inter-atom cross terms in the least-squares matrix and thus is not ideally suited for the azulene structure where there is considerable overlap of atoms.

Unobserved reflections were excluded from the analysis, with the result that 670 observational equations were available for the formation of the matrix of normal equations. Hydrogen contributions were included in the structure factor calculations but refinement of the positional parameters was carried out only for the carbon atoms. Refinement of individual isotropic temperature factors and of the scale factor was also performed. Eight cycles of least-squares refinement reduced the discrepancy between observed and calculated structure factors from 21·0% to 13·5%. The course of refinement is governed by the change in $\sum \Delta^2$ where $\Delta = (K|F_o| - |F_c|)$ and this is illustrated by Table 8. The very large R-factor quoted for Cycle 1 can be accounted for by the fact that full atoms were used instead of half-atoms in the first cycle. The magnitudes of the shifts indicated by the least-squares procedure are listed in Table 9. Since $\sum \Delta^2$ was still fairly large it was decided to refine the structure further taking account of the anisotropic thermal motion. It

Table 10

Atomic coordinates and temperature factors obtained from isotropic diagonal least-squares analysis.

Atom	x/a	y/b	z/c	X'	Y	Z'	B
A	0·140	0·037	0·327	0·587	0·223	2·514	6·219
B	0·163	0·202	0·207	0·963	1·210	1·590	5·730
C	0·068	0·116	0·050	0·460	0·694	0·384	4·536
D	0·053	0·229	-0·110	0·589	1·374	-0·842	5·235
E	-0·037	0·152	-0·268	0·127	0·908	-2·060	5·931
F	-0·139	-0·034	-0·321	-0·589	-0·205	-2·466	6·284
G	-0·168	-0·200	-0·213	-0·988	-1·199	-1·635	5·757
H	-0·111	-0·237	-0·032	-0·828	-1·421	-0·248	5·362
I	-0·009	-0·096	0·084	-0·202	-0·577	0·644	4·698
J	0·045	-0·145	0·267	-0·067	-0·868	2·055	5·925

Table 11

Deviations of carbon atoms from the mean molecular plane

$$Y = 1{\cdot}81329X' - 0{\cdot}34526Z' + 0{\cdot}00225.$$

Atom	Deviation (Å)
A	-0·012
B	-0·005
C	+0·005
D	-0·006
E	+0·017
F	-0·014
G	-0·012
H	+0·003
I	-0·005
J	+0·019

should be noted that at the end of the refinement process the average shifts in the positional parameters were fairly small.

7.1.2. Coordinates and Molecular Dimensions

The final coordinates and temperature factors of the carbon atoms after isotropic refinement are given in Table 10. The coordinates $^{x}/a$, $^{y}/b$, $^{z}/c$ are referred to the monoclinic axes with the centre of symmetry as origin, and the coordinates X', Y , Z' to orthogonal axes $\underline{a}$, $\underline{b}$ and $\underline{c}'$, $\underline{c}'$ being taken perpendicular to the $\underline{a}$ and $\underline{b}$ crystal axes, so that

$$X' = x + z\cos\beta, \quad Y' = y, \quad Z' = z\sin\beta \qquad \ldots\ldots\ldots(50)$$

These orthogonal coordinates are expressed in Ångström units.

It was found that the atomic coordinates could be fitted to an equation of the form

$$Y = AX' + BZ' + C \qquad \ldots\ldots\ldots\ldots(51)$$

the parameters A, B, C being determined by the method of least-squares to have the values 1·81329, -0·34526 and 0·00225 respectively.

The perpendicular distances of the atoms from this mean molecular plane are given in Table 11. The mean deviation is 0·009 Å and the maximum deviation is 0·019 Å. It may safely be assumed that these deviations from strict planarity are not

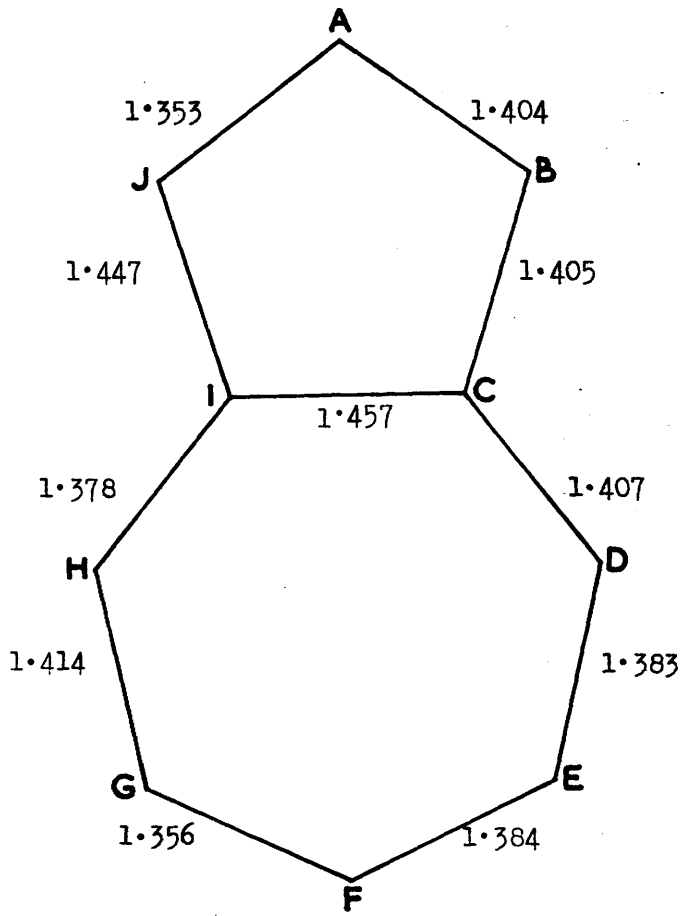


Fig. 18 **Bond lengths obtained from isotropic diagonal least-squares analysis.**

significant and that the molecule is essentially planar.

Bond lengths corresponding to the coordinates given in Table 10 were calculated and are shown in Fig. 18. It will be convenient to discuss the molecular dimensions at a later stage. However, it is appropriate to note at this stage that the interatomic distances shown in Fig. 18 cannot be considered very accurate for two reasons:-

a) anisotropic thermal vibration of the atoms was not considered in the least-squares procedure

b) interactions between strongly overlapped atoms has been neglected by using only the diagonal elements of the matrix of normal equations.

7.2. Anisotropic Diagonal Least-Squares Refinement

7.2.1. Computational Details

It was decided to refine the structure further using Rollett's programme for the English Electric "DEUCE". This programme neglects the off-diagonal terms of the matrix of normal equations but it possesses the additional facility, not available with NY XR1, of refining anisotropic thermal vibrations.

The atomic and thermal parameters listed in Table 10 were used as input data and in this refinement 70 unobserved terms were included. The latter were given values of one-half the value of their upper bound. Hydrogen contributions to the structure

factors were included but refinement was carried out only for the carbon atoms. The weighting system employed was as follows:-

$$\left.\begin{array}{l} \text{If } |F_o| \leq |F^*|, \quad \sqrt{w_2} = \dfrac{|F_o|}{|F^*|} \\ \text{If } |F_o| > |F^*|, \quad \sqrt{w_2} = \dfrac{|F^*|}{|F_o|} \end{array}\right\} \qquad \ldots\ldots\ldots\ldots(52)$$

where $|F^*| = 8\,|F_{min.}|$.

The scale factor refinement is liable to become unstable if $|F^*|$ is too small and Rollett suggests that $|F^*| = 8\,|F_{min.}|$ is a fairly safe value.

Theoretically it would be better to assign a separate weighting system to the unobserved reflections but since they represent only 19% of the total data it was decided to apply the same weighting system to them.

The programme calculates a temperature factor, T, for each atom where

$$T = \exp(-B_\theta \sin^2\theta/\lambda^2) \times 2^{-(B_{11}h^2+B_{22}k^2+B_{33}\ell^2+B_{12}hk+B_{23}k\ell+B_{31}\ell h)} \qquad \ldots\ldots\ldots\ldots(53)$$

The normal equations are solved in terms of the following sections:-

a) a (3 x 3) matrix for each atomic position

b) a (6 x 6) matrix for each atomic vibration

c) a (2 x 2) matrix for the scale factor.

Table 12

Course of refinement of anisotropic diagonal least-squares analysis

Cycle	R-factor	Cycle	R-factor
1	30·4%	7	10·4%
2	32·8%	8	9·5%
3	12·0%	9	9·5%
4	10·9%	10	10·7%
5	10·6%	11	10·1%
6	10·4%	12	9·3%

The atomic form factors used in the structure factor calculations were those of MacGillavry (1955) for carbon and of McWeeny (1952) for hydrogen. The refinement proved fairly slow, one of the reasons being that $\frac{1}{4}$-shifts only were applied at each stage. The course of refinement is shown in Table 12. The large R-factors indicated in cycles 1 and 2 are due to errors in the scale factor. After the 7th cycle it was decided to test

a) the effect of excluding the non-observed terms

b) the effect of using the weighting system $\sqrt{w_1}$, subject to the conditions:-

$$\left.\begin{aligned} &\text{if } |F_o| \leq |F^x| \;,\; \sqrt{w_1} = 1 \\ &\text{if } |F_o| > |F^x| \;,\; \sqrt{w_1} = |F^x| \,/\, |F_o| \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots (54)$$

$$\text{where } |F^x| = 8\,|F_{min.}|$$

It appears that the weighting systems $\sqrt{w_1}$ and $\sqrt{w_2}$ are equally applicable. At this stage the bond lengths were examined using the coordinates derived in Cycle 7. These proved to be most irregular and it was decided to apply shifts to atoms H and J in an effort to improve the bond lengths. The coordinates from Cycle 7, with these shifts applied, were then further refined and the final values correspond to an R-factor of 9·3%.

7.2.2. Coordinates and Molecular Dimensions

The carbon coordinates corresponding to an R-factor of 9·3%

Table 13

Atomic coordinates of carbon atoms as determined from anisotropic diagonal least-squares analysis.

Atom	x/a	y/b	z/c
A	0·140	0·037	0·328
B	0·163	0·203	0·208
C	0·068	0·115	0·050
D	0·052	0·228	-0·108
E	-0·038	0·152	-0·270
F	-0·139	-0·034	-0·322
G	-0·166	-0·200	-0·213
H	-0·110	-0·236	-0·034
I	-0·008	-0·098	0·084
J	0·045	-0·145	0·265

Table 14

Thermal parameters of carbon atoms as determined from anisotropic diagonal least-squares analysis.

Atom	B_{11}	B_{22}	B_{33}	B_{12}	B_{23}	B_{31}
A	0·035	0·065	0·034	0·001	-0·002	0·017
B	0·028	0·054	0·041	-0·005	-0·026	0·015
C	0·025	0·039	0·030	0·000	-0·002	0·015
D	0·027	0·050	0·038	0·008	0·000	0·022
E	0·032	0·063	0·033	0·006	0·004	0·020
F	0·035	0·069	0·032	0·001	-0·002	0·013
G	0·028	0·054	0·041	-0·005	-0·022	0·013
H	0·028	0·052	0·039	0·003	-0·004	0·020
I	0·026	0·041	0·030	0·009	0·008	0·011
J	0·033	0·060	0·033	0·011	0·010	0·020

Table 15

Parameter shifts indicated by anisotropic diagonal least-squares analysis

	Average initial shift	Average final shift
Δx(Å)	0·0095	0·0008
Δy(Å)	0·0102	0·0006
Δz(Å)	0·0055	0·0024
ΔB_{11}(Å^2)	0·00335	0·00011
ΔB_{22}(Å^2)	0·00962	0·00029
ΔB_{33}(Å^2)	0·00455	0·00027
ΔB_{12}(Å^2)	0·00684	0·00001
ΔB_{23}(Å^2)	0·00471	0·00054
ΔB_{31}(Å^2)	0·00133	0·00001

Table 16

Coordinates of carbon atoms referred to orthogonal axes

Atom	X'(Å)	Y(Å)	Z'(Å)
A	0·586	0·220	2·518
B	0·957	1·213	1·595
C	0·456	0·691	0·383
D	0·583	1·365	-0·830
E	0·123	0·907	-2·072
F	-0·589	-0·201	-2·476
G	-0·977	-1·200	-1·640
H	-0·811	-1·411	-0·258
I	-0·198	-0·584	0·647
J	-0·059	-0·866	2·034

Table 17

Deviations of carbon atoms from the mean molecular plane

$$Y = 1{\cdot}836X' - 0{\cdot}350Z' - 0{\cdot}003$$

Atom	Deviation (Å)
A	-0·013
B	-0·007
C	0·004
D	-0·004
E	0·019
F	-0·009
G	-0·011
H	0·004
I	-0·004
J	0·020

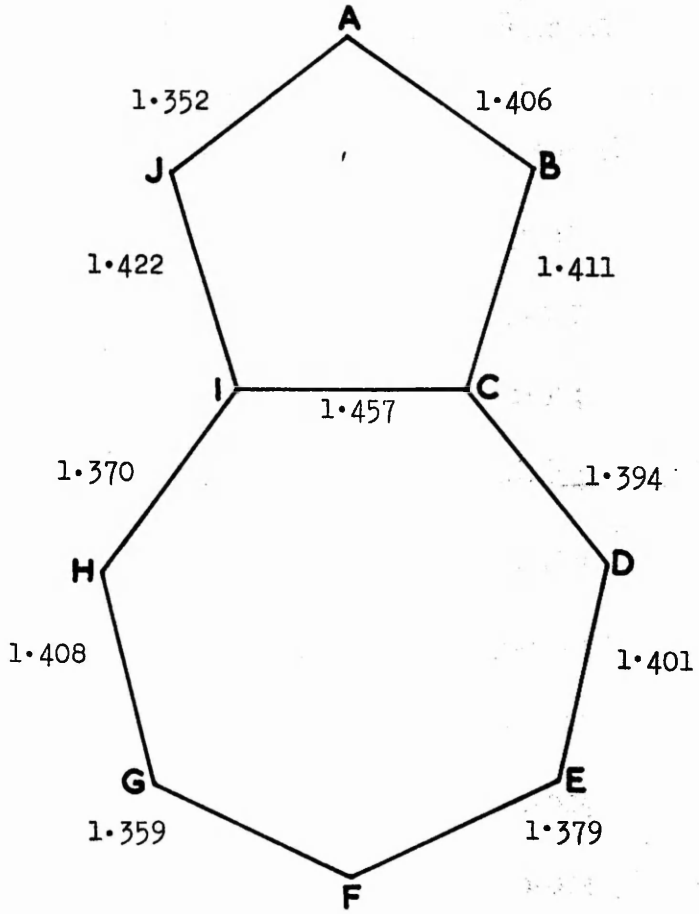


Fig. 19 Bond lengths obtained from anisotropic diagonal least-squares analysis.

are listed in Table 13 and the anisotropic thermal parameters in Table 14. The initial and final parameter shifts are indicated in Table 15. The atomic coordinates were orthogonalised by the transformation quoted in 7.1, using a programme written for "DEUCE" (see Appendix 3). The results are listed in Table 16. These orthogonal coordinates were used to compute, by the usual least-squares method, the equation of the mean molecular plane and the deviations of the carbon atoms from this plane (see Appendix 3).

The equation of the mean molecular plane is given by

$$Y = 1{\cdot}836X' - 0{\cdot}350Z' - 0{\cdot}003 \qquad \ldots\ldots\ldots\ldots(55)$$

The perpendicular distances of the atoms from this plane are given in Table 17, the average perpendicular distance being 0·010Å.

Bond lengths corresponding to the coordinates quoted in Table 16 are shown in Fig. 19.

7.3. Anisotropic Full Matrix Least-Squares Refinement

7.3.1. Computational Details

Recently a programme has become available, for use with the Oxford University "Mercury" computer, for refinement by least-squares using the full matrix of normal equations. Dr Sparks has very kindly processed the azulene data using this programme. Eight cycles of least-squares have been performed with the following characteristics:-

a) Minimisation of $\sum w \left[|F_o| - G|F_c| \right]^2$ where G is a scale factor,

Table 18

Course of refinement of full matrix least-squares analysis

Cycle	R	$\sum w\Delta^2$
1	12·8%	40·4
2	11·1%	25·7
3	6·9%	7·29
4	6·2%	5·76
5	6·2%	5·67
6	6·2%	5·60
7H	7·6%	8·08
8H	6·5%	5·56

and $w = F_o/4F_{min.}$ for $F_o \leq 4F_{min.}$

$w = 4F_{min.}/F_o$ for $F_o > 4F_{min.}$

b) Carbon scattering factors - McGillavry (1955)

Hydrogen scattering factors - McWeeny (1952)

c) Temperature factor of the form

$$\exp\left[-(B_{11}h^2+B_{22}k^2+B_{33}\ell^2+B_{12}hk+B_{13}h\ell+B_{23}k\ell)\right].$$

d) Refinement of carbon parameters and scale factor, i.e. 91 parameters.

e) Initial parameters were those obtained by the isotropic least-squares analysis (see Table 10).

After six cycles of refinement the hydrogen positions were changed to make the carbon-hydrogen bond distances equal to 1·08Å, the A - B - H(B) angle equal to the C - B - H(B) angle and H(B) was placed in the ABC plane, where H(B), etc. is the hydrogen atom attached to carbon atom B, etc. The course of refinement is indicated in Table 18 and shows a final R-factor of 6·5%. The refinement would have been terminated at Cycle 4 but for the fact that the positional parameters were still showing large shifts - as large as 0·022Å.

7.3.2. Coordinates and Molecular Dimensions

The final atomic coordinates, fractional and transformed with

Table 19

Coordinates of carbon atoms as determined from full matrix least-squares analysis.

Atom	x/a	y/b	z/c	X'(Å)	Y(Å)	Z'(Å)
A	0·144	0·053	0·330	0·616	0·319	2·535
B	0·161	0·204	0·204	0·947	1·219	1·566
C	0·068	0·118	0·050	0·459	0·706	0·384
D	0·052	0·230	-0·108	0·583	1·378	-0·832
E	-0·030	0·153	-0·263	0·173	0·918	-2·018
F	-0·134	-0·018	-0·321	-0·555	-0·110	-2·467
G	-0·170	-0·199	-0·216	-0·998	-1·194	-1·661
H	-0·110	-0·236	-0·033	-0·813	-1·414	-0·250
I	-0·008	-0·100	0·086	-0·196	-0·596	0·657
J	0·052	-0·146	0·271	-0·015	-0·875	2·078

Table 20

Coordinates of hydrogen atoms as determined from full matrix least-squares analysis

Atom	x/a	y/b	z/c
H(A)	0·199	0·078	0·464
H(B)	0·213	0·360	0·221
H(D)	0·111	0·392	-0·105
H(E)	-0·005	0·259	-0·370
H(F)	-0·195	-0·023	-0·459
H(G)	-0·254	-0·328	-0·283
H(H)	-0·150	-0·390	0·018
H(J)	0·013	-0·293	0·340

Table 21

Thermal parameters of carbon atoms as determined from full matrix least-squares analysis

Atom	B_{11}	B_{22}	B_{33}	B_{12}	B_{23}	B_{13}
A	0·0189	0·0490	0·0290	-0·0074	-0·0158	0·0121
B	0·0204	0·0265	0·0296	-0·0069	-0·0109	0·0058
C	0·0149	0·0229	0·0224	-0·0009	0·0002	0·0108
D	0·0198	0·0268	0·0250	0·0004	0·0060	0·0147
E	0·0199	0·0607	0·0244	0·0032	0·0097	0·0121
F	0·0286	0·0504	0·0172	0·0209	-0·0010	0·0039
G	0·0178	0·0574	0·0218	0·0018	-0·0222	0·0070
H	0·0182	0·0271	0·0288	-0·0011	-0·0046	0·0120
I	0·0164	0·0256	0·0220	0·0019	0·0027	0·0119
J	0·0205	0·0332	0·0204	0·0114	0·0081	0·0088

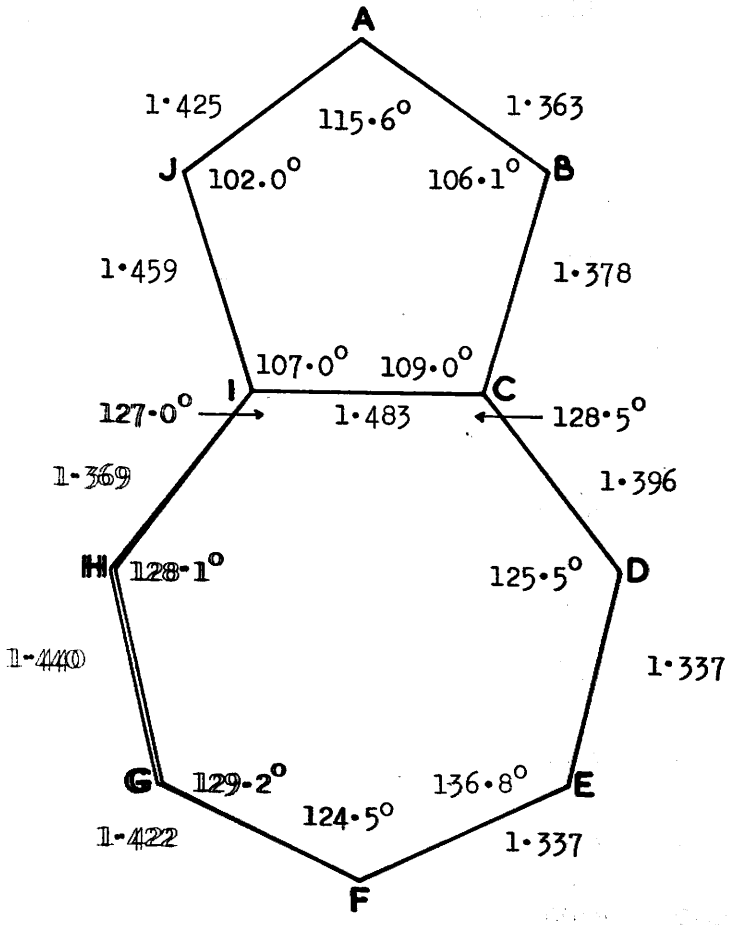


Fig. 20 **Final bond lengths and bond angles.**

Table 22.

Deviations of carbon atoms from mean molecular plane.

(Deviations in Å)

Atom	Deviations (a)	Deviations (b)
A	-0·038	-0·037
B	-0·016	-0·013
C	0·000	0·001
D	-0·009	-0·007
E	0·050	0·051
F	-0·022	-0·022
G	-0·027	-0·029
H	0·004	0·001
I	0·002	0·001
J	0·055	0·054

respect to orthogonal axes, are given, for carbon atoms, in Table 19, and for hydrogen atoms in Table 20. The anisotropic temperature parameters for the carbon atoms are presented in Table 21. The analysis of the latter in terms of rigid-body vibrations will be discussed later. The bond lengths and bond angles for the azulene molecule are shown in Fig. 20, these dimensions having been computed from the atomic coordinates listed in Table 19.

By the usual least-squares method the equation of the mean molecular plane was determined to be

$$Y = 1{\cdot}835X' - 0{\cdot}351Z' - 0{\cdot}002 \qquad \ldots\ldots\ldots\ldots(56)$$

The perpendicular distances of the carbon atoms from this plane are listed in Table 22, in the column headed 'Deviation (a)'. The average deviation of the atoms from the mean molecular plane is 0·022Å.

Recently another method of determining the 'best' plane through a set of atoms has been developed (Schomaker et al, 1959). In this case the weighted sum of the squares of the distances of the atoms from the plane is minimised. This problem reduces to that of finding the eigenvalues of a matrix of order 3. The characteristic equation has roots, $\lambda^{(1)} \ll \lambda^{(2)} \leq \lambda^{(3)}$, corresponding to the 'best' plane, an intermediate plane, and a 'worst' plane, all at right angles to one another. Because of the smallness of the eigenvalue $\lambda^{(1)}$, the value of $\lambda^{(1)}$ can be

conveniently determined by the iterative method of matrix powering (Frazer et al, 1938). A programme has been written for this computation and details are given in Appendix 3. The deviations of the carbon atoms, determined by Schomaker's method, are listed in Table 22, in the column headed 'Deviation (b)'. The mean deviation in this case has a value of 0.030Å. The final values of F_o and F_c, corresponding to Cycle 8H, are listed in Appendix 2.

8. Analysis of Thermal Motion

8.1. Theory

In many organic crystals the molecular vibrations are large compared with the atomic movements, due to the relatively weak van der Waals attraction. Accordingly the thermal motion can be described in terms of rigid-body vibrations of the molecule as a whole. A detailed treatment of the problem has been given by Cruikshank (1956) in which the thermal displacements are expressed in terms of an ellipsoidal distribution.

The vibrations of an atom in an anisotropic harmonic potential field may be represented by a symmetric tensor, U, of the form

$$U = \begin{vmatrix} U_{11} & U_{12} & U_{13} \\ U_{12} & U_{22} & U_{23} \\ U_{13} & U_{23} & U_{33} \end{vmatrix} \qquad \text{.....(57)}$$

The mean square amplitude of vibration $\overline{u^2}$ in the direction of

a unit vector $\vec{l}$, with components l_i, is then

$$\overline{u^2} = \sum_{i=1}^{3} \sum_{j=1}^{3} U_{ij} l_i l_j \qquad \text{.....(58)}$$

It is convenient to consider the system referred to a set of molecular axes, the transformation from the orthogonal set to the molecular set being effected by the relation

$$\begin{pmatrix} X_m \\ Y_m \\ Z_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_j \\ y_j \\ z_j \end{pmatrix} \qquad \text{....(59)}$$

where the a_{ij} are the direction-cosines of the molecular axes with respect to the orthogonal axes. The anisotropic temperature parameters, B_{ij}, may be transformed from monoclinic crystal axes to orthogonal axes as shown by Rollett and Davies (1955), and finally transformed, by the usual tensor rules (Nye, 1957) to molecular axes, using the expression

$$U_{ij} = a_{ik} a_{kj} b_{ij} \qquad \text{.....(60)}$$

where the b_{ij} are the values of the B_{ij} transformed with respect to orthogonal axes. In terms of rigid-body vibrations, the motion of the molecule can be expressed as two symmetric tensors T and ω , each with six independent components. T represents

the translational vibrations of the mass centre and ω the angular oscillations about the centre.

By a least-squares procedure it is possible to express the T_{ij} and ω_{ij} in terms of U_{ij}. Conversely, the U tensors for each atom may be calculated from the T and ω tensors.

The accuracy of the determination of the T_{ij} and ω_{ij} is estimated by the formula

$$\sigma^2(A_p) = C_{pp}^{-1} \sigma^2(U) \quad \ldots\ldots\ldots\ldots\ldots\ldots(61)$$

where $\sigma^2(A_p)$ is the variance of one of the T_{ij} or ω_{ij}, C_{pp}^{-1} is the appropriate diagonal element of the inverse matrix and $\sigma^2(U)$ is given by

$$\sigma^2(U) = \sum_n (U_n^{obs} - U_n^{calc})^2/t \quad \ldots\ldots\ldots\ldots(62)$$

where t is the difference between the total number of U_{ij}^{obs} and the number of parameters determined.

8.2. Application to Azulene

For azulene, molecular axes were chosen as in Fig. 21, with OZ_m perpendicular to both OX_m and OY_m.

Table 23

Observed and Calculated U_{ij}.

(Values in 10^{-2} Å^2)

	U_{11}		U_{22}		U_{33}	
Atom	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
A	8·59	6·54	8·12	10·42	5·70	6·19
B	8·44	7·78	5·16	7·43	6·14	5·57
C	6·95	7·02	3·89	4·81	3·68	3·79
D	8·07	8·76	4·39	4·83	4·61	4·83
E	7·88	7·84	9·32	7·46	6·20	5·50
F	5·46	6·54	10·94	9·98	6·64	5·97
G	6·26	7·74	10·56	7·74	5·16	5·71
H	8·70	7·75	4·99	4·82	4·47	4·25
I	7·00	6·99	4·42	4·82	3·78	3·75
J	6·57	7·66	6·45	7·74	4·70	5·53

Table 24

Observed and Calculated U_{ij}.

(Values in 10^{-2} Å^2)

	U_{12}		U_{23}		U_{13}	
Atom	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
A	1·65	0·35	1·97	0·90	0·90	0·54
B	1·78	2·01	0·37	0·46	0·21	0·56
C	0·12	0·46	0·40	0·36	0·27	0·47
D	0·69	0·65	0·41	0·45	0·44	0·21
E	1·46	1·82	2·64	0·84	0·22	0·22
F	1·18	0·10	1·11	0·90	0·57	0·43
G	1·70	2·07	1·72	0·51	1·07	0·54
H	0·75	0·67	0·21	0·34	0·39	0·43
I	0·35	0·21	0·40	0·42	0·33	0·41
J	1·29	1·77	0·27	0·80	0·18	0·16

Table 25

Values of T_{ij} and ω_{ij}.

(Values of T_{ij} in 10^{-2} Å^2 and of ω_{ij} in deg.^{2})

$$T = \begin{pmatrix} 6\cdot53 & 0\cdot12 & 0\cdot48 \\ & 4\cdot57 & 0\cdot37 \\ & & 3\cdot37 \end{pmatrix} \qquad \omega = \begin{pmatrix} 17\cdot10 & 0\cdot45 & 2\cdot46 \\ & 13\cdot40 & \overline{2\cdot63} \\ & & 27\cdot83 \end{pmatrix}$$

Table 26

Values of $\sigma(T_{ij})$ and $\sigma(\omega_{ij})$.

(Values of $\sigma(T_{ij}$ in 10^{-2} Å^2 and of $\sigma(\omega_{ij})$ in deg.^{2})

$$\sigma(T) = \begin{pmatrix} 0\cdot36 & 0\cdot34 & 0\cdot40 \\ & 0\cdot48 & 0\cdot48 \\ & & 0\cdot96 \end{pmatrix} \qquad \sigma(\omega) = \begin{pmatrix} 4\cdot10 & 1\cdot94 & 2\cdot57 \\ & 2\cdot32 & 1\cdot99 \\ & & 1\cdot95 \end{pmatrix}$$

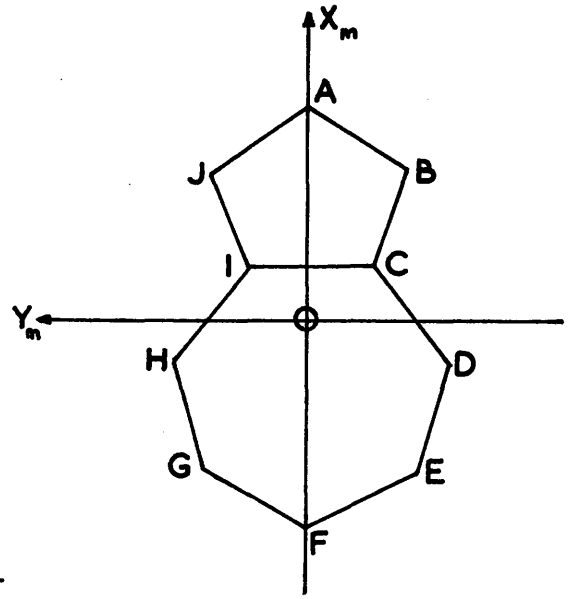


Fig. 21.

The values of the six independent U_{ij}'s for each carbon atom are shown in Tables 23 and 24, in the columns headed 'obs.'. The T and ω tensors for azulene are shown in Table 25. Taking square roots of the diagonal terms we find that the root mean square amplitudes of translational vibration in the direction of the molecular axes are:-

along	OX_m	0·26 Å
along	OY_m	0·21 Å
along	OZ_m	0·18 Å

and the corresponding root mean square amplitudes of angular oscillation are:-

about	OX_m	$4{\cdot}1^\circ$
about	OY_m	$3{\cdot}7^\circ$
about	OZ_m	$5{\cdot}3^\circ$

From the T and ω tensors, the U tensors for each atom have been calculated and the values of U_{ij} are listed in Tables 23 and 24, in the columns headed 'Calc.'. The root mean square difference between the $U_{ij}^{obs.}$ and the $U_{ij}^{calc.}$ is 0·0105 Å^2, corresponding to an estimated standard deviation for the $U_{ij}^{obs.}$ of 0·0107 Å^2. The corresponding e.s.d.'s of the T_{ij} and ω_{ij} are shown in Table 26.

The values of the T and ω tensors are rather similar to those for naphthalene and it is seen that the greatest amplitude of translational vibration is along the long axis of the molecule as is the case with naphthalene (Cruickshank, 1957a) and anthracene (Cruickshank, 1957b). It is interesting to note that in the case of naphthalene the corresponding root mean square amplitudes of translational vibration are 0·22, 0·20, 0·19 Å and of the angular oscillation 4·4°, 3·7° and 4·2°. Unfortunately no Raman data for azulene are available for comparison with the amplitudes of angular oscillation.

9. Discussion of Results

In the past the accuracy of an X-ray structure analysis was generally estimated by consideration of the consistency of the

Table 27

Bond lengths and standard deviations corresponding to Fig. 20.

Bond	Bond Length (Å)	s.d. (Å)
AB	1·363	0·027
BC	1·378	0·019
CD	1·395	0·005
CI	1·483	0·004
DE	1·337	0·034
EF	1·337	0·035
FG	1·422	0·026
GH	1·440	0·021
HI	1·369	0·007
IJ	1·459	0·024
JA	1·425	0·032
AF'	0·228	0·052
BG'	0·111	0·042
EJ'	0·175	0·064

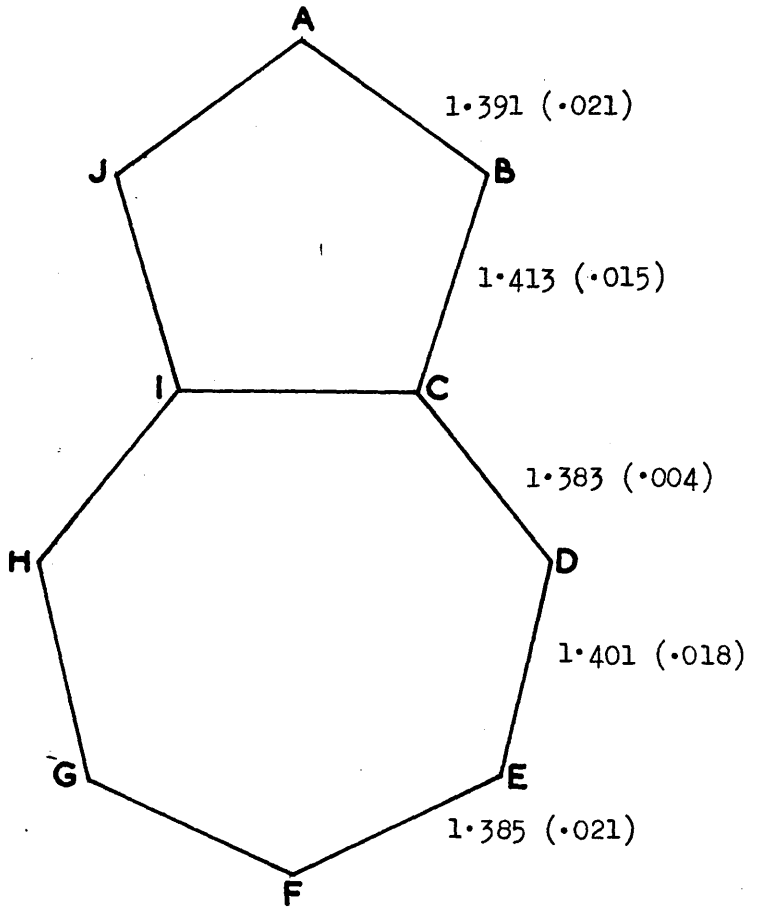


Fig. 22 Weighted average bond lengths for equivalent bonds, with standard deviations in parentheses.

results or by the effect of variation of the parameters on the agreement index. More recently a quantitative theory of accuracy has been developed, based on the standard deviations of atomic positions (Cruickshank, 1949, 1954; Cruickshank and Rollett, 1953).

A rigorous discussion of the accuracy with which the bond lengths of azulene have been determined is very difficult since, due to the disordered structure, the overlap of atoms is very marked. The actual distances of separation for the atom-pairs are as follows:-

A - F'	0·228 Å
B - G'	0·111 Å
C - H'	0·803 Å
D - I'	0·890 Å
E - J'	0·175 Å

Sparks' programme for full matrix least-squares analysis evaluates the standard deviations of the bond lengths, using the full inverse matrix and including all variances and covariances. The standard deviations of the bond lengths shown in Fig. 20 are listed in Table 27. Assuming that equivalent bond lengths are uncorrelated, weighted averages for the equivalent bonds have been evaluated and are illustrated, with standard deviations in parentheses, in Fig. 22. These appear to conform roughly to the usual benzenoid-type bond lengths with the transannular C-C bond much longer than the other

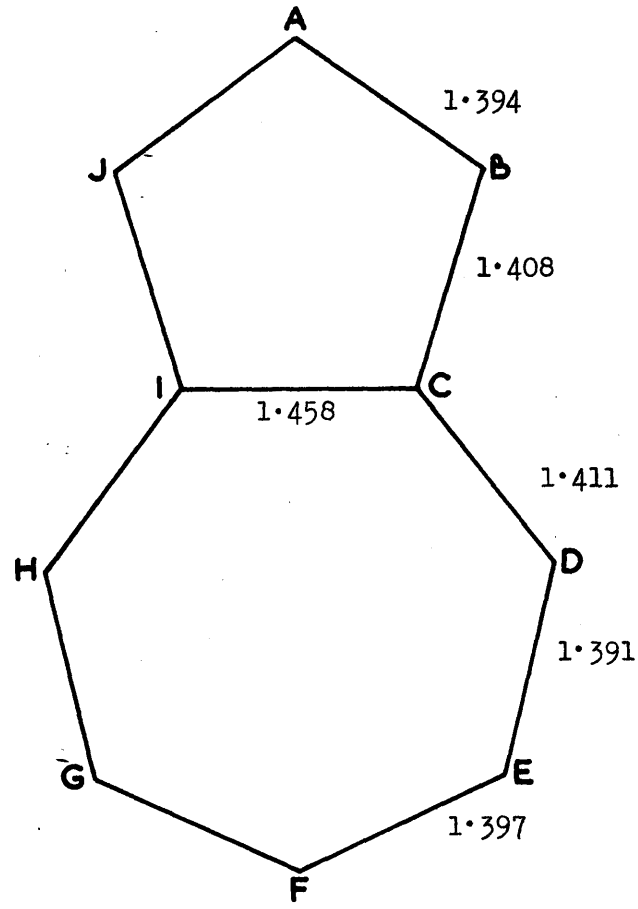
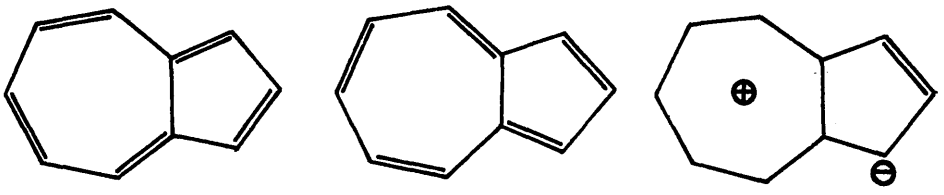


Fig. 23 Bond lengths calculated by the method of molecular orbitals (R.D. Brown, 1948).

bonds in the molecule. However, it is significant that the χ^2 value for the weighted set corresponds, with five degrees of freedom, to a probability value of less than 0·01. Recently results have become available of X-ray work done on heterocyclic azulenes (Sasada, 1958, 1959; Tamura et al, 1958) and the only results which can reasonably be compared with those of azulene are the bond lengths of 2-amino azulene (Takaki et al, 1959). In this compound the central C-C bond has a length of 1·52 Å and the other C-C bond lengths range from 1·37 Å to 1·41 Å.

Theoretical studies of the azulene molecule by the method of molecular orbitals have been made by, among others, Estelles and Alonso (1952) and Brown (1948). The results obtained by the latter are shown in Fig. 23. These results are in agreement with the values to be expected on the basis of the contributing Kekulé structures.



(a) (b) (c)

(a) and (b) are the most important structures contributing to the azulene hybrid, with a small contribution from the dipolar form(c) to account for the dipole moment, μ = 1·0 D, observed

by Wheland and Mann (1949). Thus it is to be expected that all bonds will have approximately double-bond character except the bond common to both rings which will have predominantly single-bond character. In a recent Chemical Society symposium (Chemical Society Symposia, 1958) Dewar has predicted that the central bond should have a length of 1·47 ± 0·01 Å, the value to be expected for a single bond between sp^2 hybridised carbon atoms. As noted earlier, atoms C and I have been determined with the greatest accuracy in our analysis and thus we feel certain that the predictions regarding the length of the trans-annular bond are correct. Intermolecular distances have been calculated and the values are those which would be expected, ranging from 3·6 to 3·8 Å.

In view of the inaccuracies inherent in this structure determination no significance can be placed on the deviations of the atoms from strict planarity. Nevertheless, it is of interest to note that the deviations of atoms C, D, H, I are fairly small, these being the atoms involving least overlap and hence most accurately determined. It should be noted, however, that all theoretical studies have been based on the assumption that the azulene molecule is planar.

The highly unsymmetrical nature found for the molecule has led us to consider possible alternative solutions to the azulene structure. In spite of the fact that the R-factor at the final

stage is 6·5%, the coordinate shifts amount to as much as 0·02 Å.

One possibility is that there is no exact symmetry between the two possible positions and that the sites are equally occupied. This would imply that the (0k0) absences are indeed accidental and that the molecules are free to move independently relative to one another in the space group Pa.

Alternatively, but less likely, the above situation could hold with the modification that the sites are unequally occupied, i.e. with a greater proportion of the molecules pointing, say, 'north' than pointing 'south'.

Under the present circumstances it is important to bear in mind the observation by Bernal (1956) that the morphology of azulene crystals grown by sublimation does not indicate a two-fold axis. In addition the results of Takeuchi and Pepinsky (1956) might be re-examined, preferably with full three-dimensional data. Unfortunately it is not known by which process they obtained the azulene crystals.

At the moment work is being done (Ross, private communication) on azulene with reference to crystal field spectral splittings. Measurements at $4^{\circ}K$ on sublimation flakes appear to yield incompatible results. Visible absorption studies indicate disorder in the crystal whereas the polarised infra-red spectrum apparently indicates an ordered crystal. This work, however, is still incomplete and it will be interesting to learn the results when more

accurate measurements have been made.

Thus it would appear that the structure of azulene is in rather an uncertain position at the moment. However, we feel that our results must be a fairly good approximation to the truth and it is hoped to complete these studies in terms of the electron density distribution in the near future.

References

Albrecht, G. (1939). Rev. Sci. Instr. 10, 221.

Arigoni, D., Viterbo, R., Dünnenberger, M., Jeger, O., and Ruzicka, L. (1954). Helv. Chim. Acta. 37, 2306.

Arnott, S. and Robertson, J.M. (1959). Acta Cryst. 12, 75.

Barton, D.H.R. (1953) J. Chem. Soc. p.1027.

Barton, D.H.R. and Elad, D. (1956). J. Chem. Soc. p.2085.

Barton, D.H.R., McGhie, J.F., Pradhan, M.K. and Knight, S.A. (1954 Chem. and Ind. p.1325.

Barton, D.H.R., McGhie, J.F., Pradhan, M.K. and Knight, S.A. (1955) J. Chem. Soc. p.876.

Beevers, C.A. and Robertson, J.H. (1950). Acta Cryst. 3, 164.

Bernal, J.D. (1956). Nature (London) 178, 40.

Bernays (1841). Annalen. 40, 317.

Booth, A.D. (1946). Trans. Faraday Soc. 42, 444.

Booth, A.D. (1946). Trans. Faraday Soc. 42, 617.

Booth, A.D. (1947). Nature (London) 160, 196.

Booth, A.D. (1948). Nature (London) 161, 765.

Booth, A.D. (1949). Proc. Roy. Soc. A197, 336.

Brachvogel, L. (1952). Arch. Pharmaz. 285, 57.

Bragg, W.H. (1915). Trans. Roy. Soc. (London) A215, 253.

Bragg, W.L. (1929). Proc. Roy. Soc. (London) A123, 537.

Brindley, G.W. and Wood, R.G. (1929). Phil. Mag. 7, 616.

Brindley, G.W. and Wood, R.G. (1929). Phil. Mag. 7, 619.

Brown, R.D. (1948). Trans. Faraday Soc. 44, 984.

Buerger, M.J. (1950). Acta Cryst. 3, 87.

Buerger, M.J. (1951). Acta Cryst. 4, 531.

Carlisle, C.H. and Crowfoot, D. (1945). Proc. Roy. Soc. A184, 64.

Chemical Society Symposia. Special Publication No. 12. London: The Chemical Society. 1958.

Clastre, J. and Gay, R. (1950a). C.R. Acad. Sci., Paris. 230, 1976.

Clastre, J. and Gay, R. (1950b). J. Phys. Radium 11, 75.

Clews, C.J.B. and Cochran, W. (1948). Acta Cryst. 1, 4.

Cochran, W. (1951). Acta Cryst. 4, 408.

Crowfoot, D., Bunn, C.W., Rogers-Low, B.W. and Turner-Jones, A. (1949). The Chemistry of Penicillin, Princeton: Princeton Universtiy Press.

Cruickshank, D.W.J. (1949). Acta Cryst. 2, 65.

Cruickshank, D.W.J. (1954). Acta Cryst. 7, 519.

Cruickshank, D.W.J. (1956). Acta Cryst. 9, 757.

Cruickshank, D.W.J. (1957a). Acta Cryst. 10, 504.

Cruickshank, D.W.J. (1957b). Acta Cryst. 10, 470.

Cruickshank, D.W.J. and Rollett, J.S. (1953). Acta Cryst. 6, 705.

Curtis, R.G., Fridrichsons, J. and Mathieson, A.M. (1952). Nature (London) 170, 321.

Dean, F.M. and Geissman, T.A. (1958). J. Org. Chem. 23, 596.

Emerson, O.H. (1948). J. Amer. Chem. Soc. 70, 545.

Emerson, O.H. (1951). J. Amer. Chem. Soc. 73, 2621.

Estelles, I. and Alonso, J.I.F. (1952). Anales real Soc. españ. Fís. Quím. 48B, 115.

Frazer, R.A., Duncan, W.J. and Collar, A.R. (1938). Elementary Matrices and Some Applications to Dynamics and Differential Equations. Cambridge: The University Press.

Fujita, A. and Hirose, Y. (1954). J. Pharm. Soc. Japan 74, 365.

Fujita, A., and Hirose, Y. (1956). J. Pharm. Soc. Japan 76, 129.

Garrido, J. (1950a). C.R. Acad. Sci., Paris 230, 1878.

Garrido, J. (1950b). C.R. Acad. Sci., Paris 231, 297.

Grenville-Wells, H.J. and Abrahams, S.C. (1952). Rev. Sci. Instr. 23, 328.

Günthard, H.H. (1949). Thèse Doc. Sci. tech. Zurich, Neographic, in 8°.

Günthard, H.H. (1955). Helv. Chim. Acta 38, 1918.

Günthard, H.H., Plattner, Pl.A. and Brandenberger, E. (1948). Experientia 4, 425.

Harker, D. (1936). J. Chem. Phys. 4, 381.

Hartree, D.R. (1928). Proc. Camb. Phil. Soc. 24, 89, 111.

Hendricks, S.B. (1933). Z. Krist. 84, 85.

Hoerni, J.A. and Ibers, J.A. (1954). Acta Cryst. 7, 744.

Howells, E.R., Phillip, D.C. and Rogers, D. (1950). Acta Cryst. 3, 210.

Hughes, E.W. (1941). J. Amer. Chem. Soc. 63, 1737.

International Tables for X-Ray Crystallography (1952). Vol. I Birmingham: Kynoch Press.

James, R.W. and Brindley, G.W. (1931). Z. Krist. 78, 470.

James, R.A. and Brindley, G.W. (1932). Phil. Mag. 12, 81.

Jones, F.T. and Palmer, K.J. (1949). J. Amer. Chem. Soc. 71, 1935.

Klug, A. (1947). Nature (London) 160, 570.

Koller, G. and Czerny, H. (1936). Mh. Chem. 67, 248.

Lipson, H. and Cochran, W. (1953). The Determination of Crystal Structures. London: G. Bell & Sons Ltd.

Lonsdale, K. (1936). Structure Factor Tables. London: Bell.

Luzzati, V. (1953). Acta Cryst. 6, 142.

MacGillavry, C.H., Berghuis, J., Haanappel, IJ.M., Potters, M., Loopstra, B.O. and Veenendaal, A.L. (1955). Acta Cryst. 8, 478.

McLachlan, D. (1951). Proc. Nat. Acad. Sci. 37, 115.

McWeeny, R. (1951). Acta Cryst. 4, 513.

McWeeny, R. (1952). Acta Cryst. 5, 463.

Martin, A.J.P. (1931). Min. Mag. 22, 519.

Maurice, M.E. (1930). Proc. Camb. Phil. Soc. 24, 491.

Melera, A., Schaffner, K., Arigoni, D. and Jeger, O. (1957). Helv. Chim. Acta 40, 1420.

Misch, L. and van der Wyk, A.J. (1937). Compt. rend. soc. phys. hist. nat. Genève 54, 106.

Nye, J.F. (1957). Physical Properties of Crystals. Oxford: Clarendon Press.

Patterson, A.L. (1935). Z. Krist. 90, 517.

Pfau, A. St. and Plattner, Pl.A. (1936). Helv. Chim. Acta 19, 858.

Qurashi, M.M. (1949). Acta Cryst. 2, 404.

Robertson, J.M. (1935). J. Chem. Soc. p.615.

Robertson, J.M. (1936). J. Chem. Soc. p.1195.

Robertson, J.M. (1943). J. Sci. Instr. 20, 175.

Robertson, J.M. (1953). Organic Crystals and Molecules. Ithaca, New York: Cornell University Press.

Robertson, J.M. (1954). Acta Cryst. 7, 817.

Robertson, J.M. (1955). Acta Cryst. 8, 286.

Robertson, J.M. and Shearer, H.M.M. (1956). Nature (London) 177, 885.

Robertson, J.M., Shearer, H.M.M., Sim, G.A., and Watson, D.G. (1958) Nature (London) 182, 177.

Robertson, J.M. and Todd, G. (1953). Chem. and Ind. p.437.

Robertson, J.M. and Woodward, I. (1937). J. Chem. Soc. p.219.

Robertson, J.M. and Woodward, I. (1940). J. Chem. Soc. p.36.

Rollett, J.S. and Davies, D.R. (1955). Acta Cryst. 8, 125.

Rossmann, M.G. (1956). Acta Cryst. 9, 819.

Sasada, Y. (1959). Bull. Chem. Soc. Japan. 32, 165.

Sasada, Y. (1959). Bull. Chem. Soc. Japan. 32, 171.

Schomaker, V., Waser, J., Marsh, R.E. and Bergman, G. (1959). Acta Cryst. 12, 600.

Shoemaker, D.P., Donohue, J., Schomaker, V. and Corey, R.B. (1950). J. Amer. Chem. Soc. 72, 2328.

Sim, G.A. (1959). Acta Cryst. 12, 813.

Sondheimer, F., Meisels, A. and Kincl, F.A. (1959). J. Org. Chem. 24, 870.

Takaki, Y., Sasada, Y. and Nitta, I. (1959). J. Phys. Soc. Japan 14, 771.

Takeuchi, Y. and Pepinsky, R. (1956). Science 124, 126.

Tamura, C., Sasada, Y. and Nitta, I. (1959). Bull. Chem. Soc. Japan 32, 458

Tunell, G. (1939). Am. Min. 24, 448.

Vand, V. (1948). Nature (London) 161, 600.

Vand, V. (1949). Nature (London) 163, 129.

Vand, V. (1951). Acta Cryst. 4, 285.

Waser, J. (1951). Rev. Sci. Instr. 22, 567.

Watson, G.N. (1922). A Treatise on the Theory of Bessel Functions. Cambridge: University Press.

Wheland, G.W. and Mann, D.E. (1949). J. Chem. Phys. 17, 264.

Whittaker, E.T. and Robinson, G. (1944). The Calculus of Observations 4th ed. Glasgow: Blackie.

Wilson, A.J.C. (1949). Acta Cryst. 2, 318.

Wilson, A.J.C. (1951). Research p.141.

Woolfson, M.M. (1956). Acta Cryst. 9, 804.

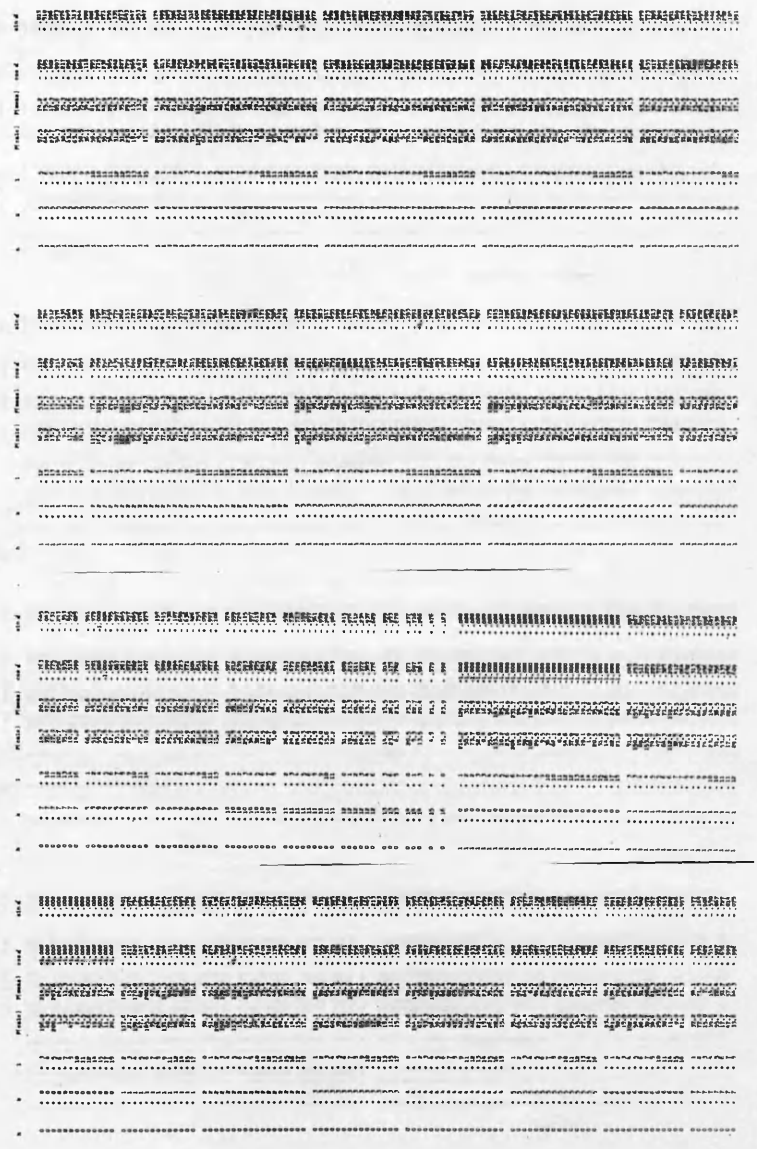
Wooster, W.A. (1949). Crystal Physics. Cambridge: University Press.

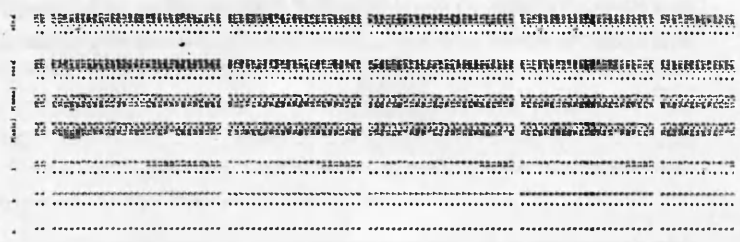
Wrinch, D.M. (1939). Phil. Mag. 27, 98.

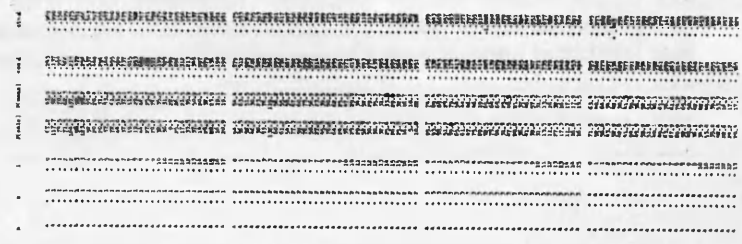
Yü, S.H. (1942). Nature (London) 149, 638.

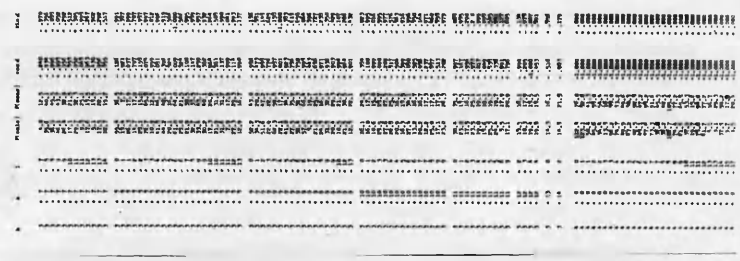
Appendix 1

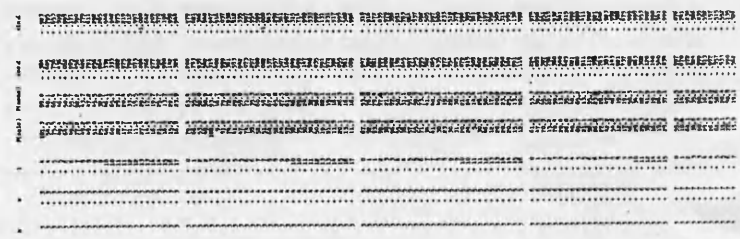
Observed and calculated structure factors for epi-limonol iodoacetate. Unobserved terms are omitted.

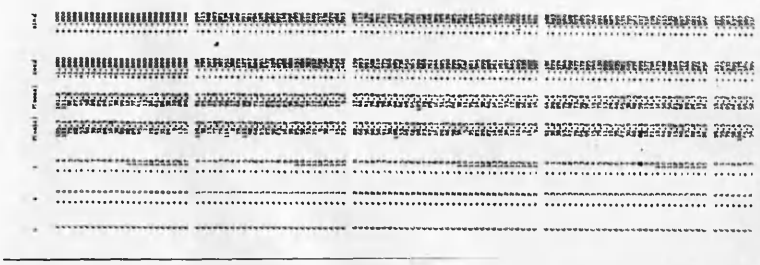


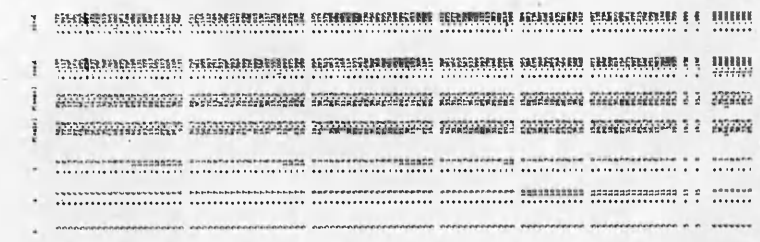


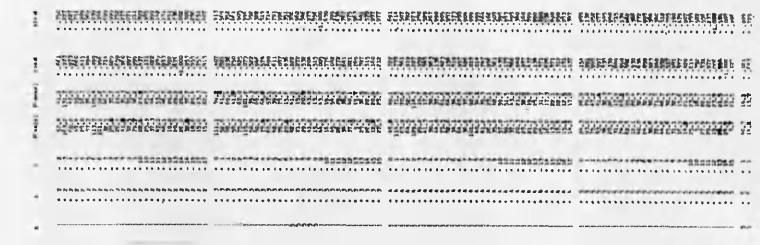


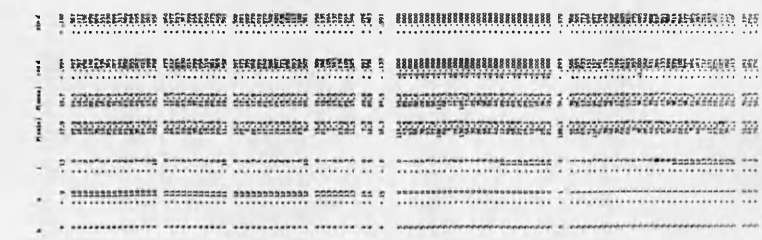


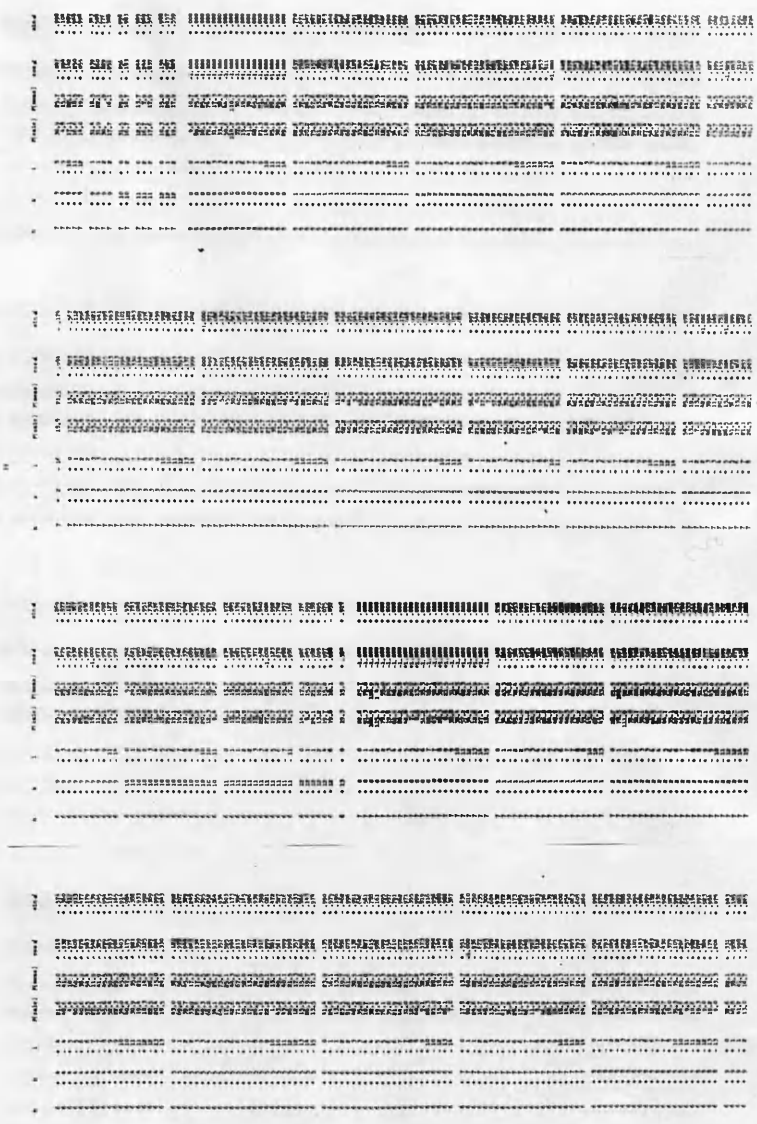


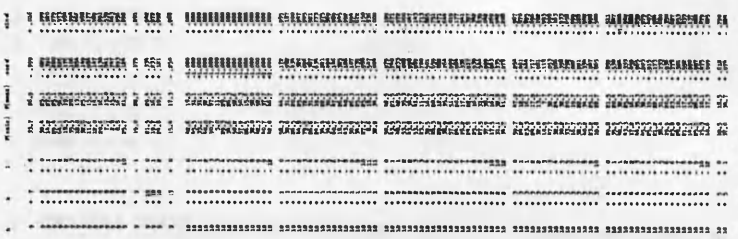


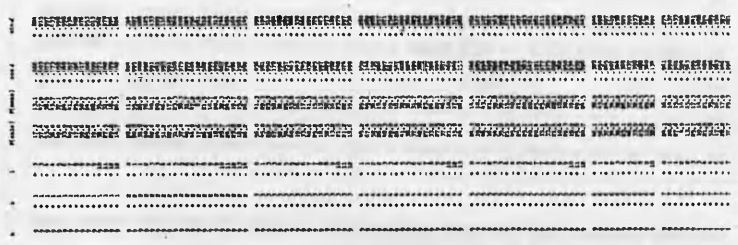


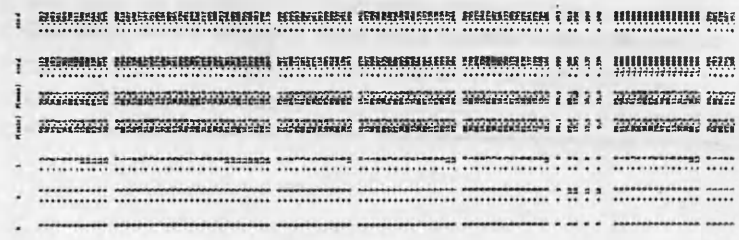


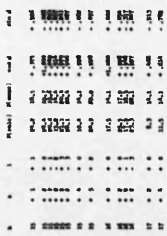


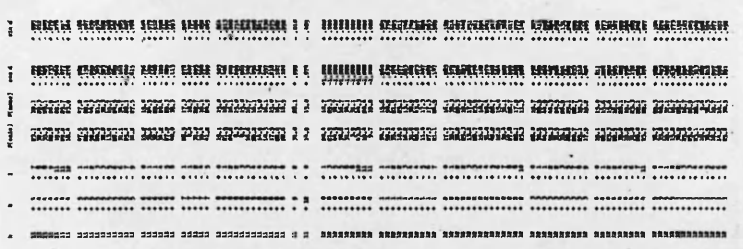












Appendix 2.

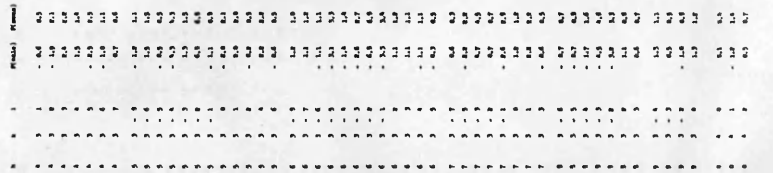
Observed and calculated structure factors for azulene.

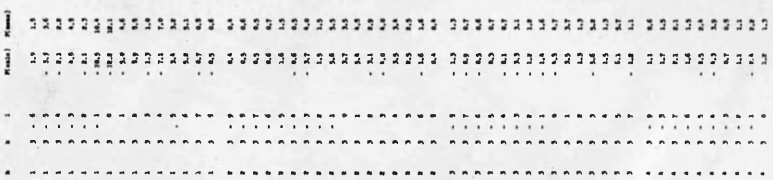
Unobserved terms are omitted.

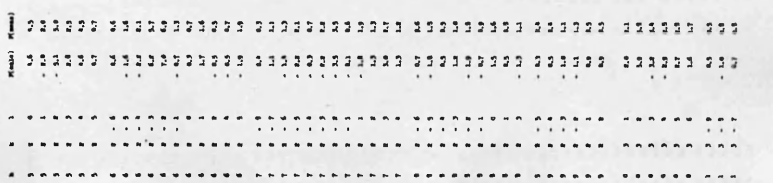
[illegible]

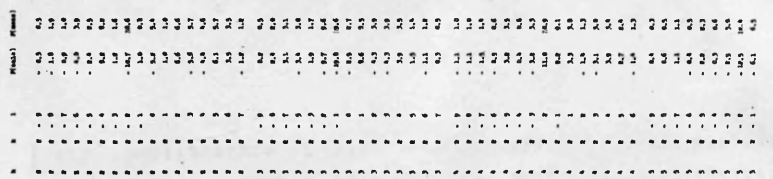
[illegible]

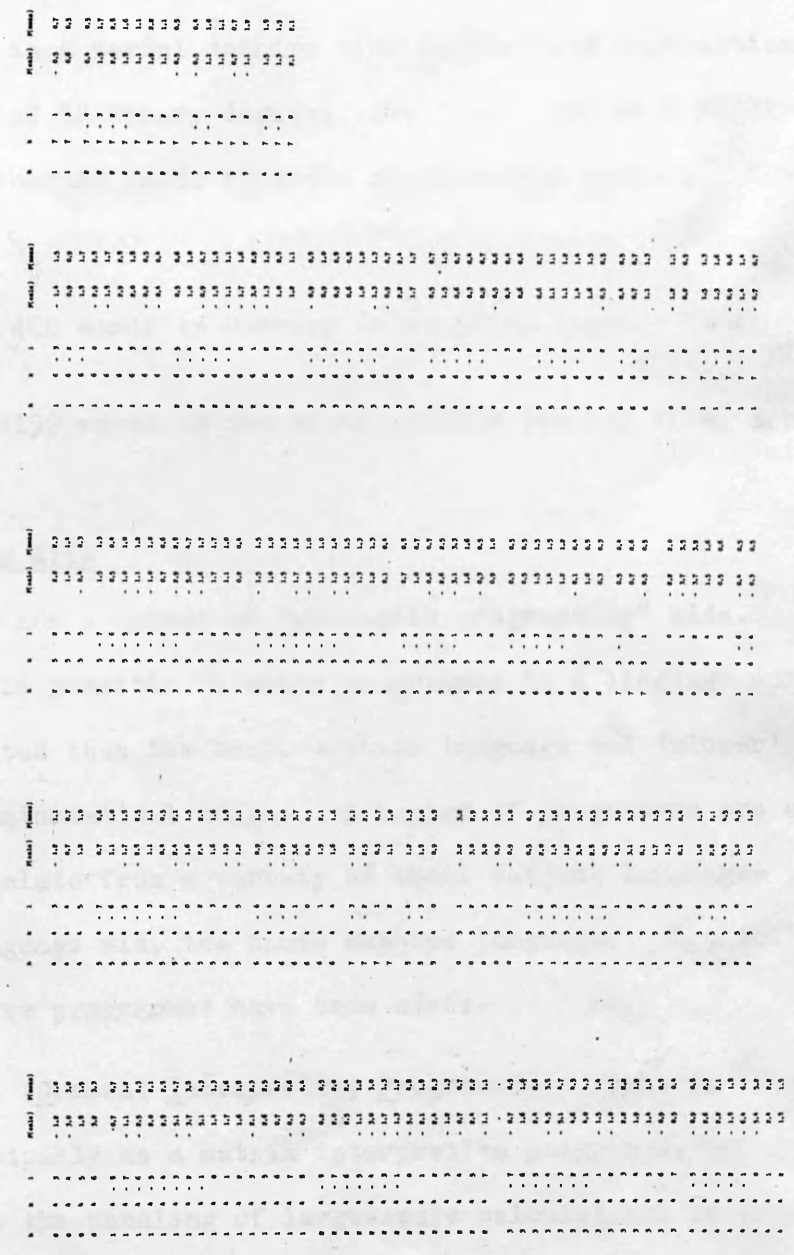
[illegible]











Appendix 3

3.1. Synopsis of 'DEUCE'

DEUCE is a serial machine with numbers and instructions consisting of 32 binary digits; the digit rate is 1 million per second. Punched cards form the input-output medium. The store comprises

(1) 402 words in mercury delay lines (access time, 32-1024 microseconds)

and (2) 8192 words on the magnetic drum (access time, approx. 13-48 milliseconds)

Programming Aids

There are a number of "automatic programming" aids. Using these, it is possible to write programmes in a language more sophisticated than the basic machine language and 'closer' to accepted mathematical usage. A number of programmes are available which translate from a variety of these subject languages to the object language viz, the basic machine language. The following interpretive programmes have been used:-

1. G.I.P. (General Interpretive Programme). This programme is used principally as a matrix interpretive programme, being well adapted to the handling of large-scale calculations on arrays of numbers.

2. Alphacode. This programme handles protracted calculations involving single variables; it is best used on so-called "one-off" jobs e.g., exploratory calculations in research studies.

3.2. Programme for Transformation of Atomic Coordinates

Description The programme transforms atomic coordinates referred to monoclinic crystal axes to coordinates referred to a set of orthogonal axes.

Input The input consists of

(i) the number of atoms, ν ($\nu \leq 100$)

(ii) a, b, c, $\cos\beta$, $\sin\beta$

(iii) ν sets of $^{x}/_{a}$, $^{y}/_{b}$, $^{z}/_{c}$.

Output The atomic coordinates X', Y, Z' and x, y, z.

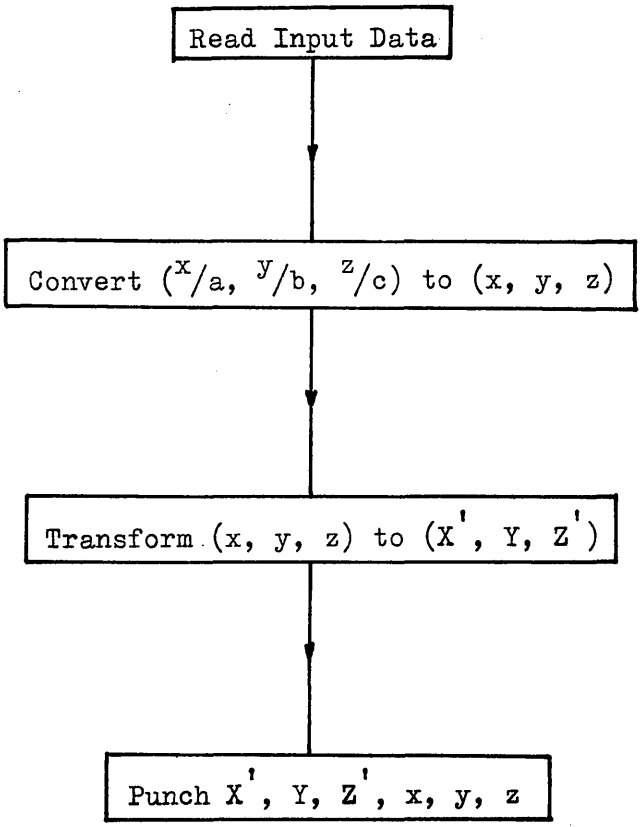
Time Approximately 2 minutes for 25 atoms.

System Alphacode Mark I.

Method Let the crystal axes be a, b, c and the orthogonal axes a', b and c', c' being taken perpendicular to the a and b crystal axes. If the coordinates (in Å) referred to the monoclinic axes are denoted by x, y, z and those with respect to the orthogonal axes by X', Y, Z', then

$$\left.\begin{aligned} X' &= x + z\cos\beta \\ Y' &= y \\ Z' &= z\sin\beta \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (63)$$

Flow Diagram



3.3. Programme for Calculation of the Equation of the Mean Molecular Plane and the Deviations of the Atoms from this Plane.

Description This programme calculates the equation of the 'best' plane through the atoms of a molecule which is assumed to be almost planar and computes the perpendicular distance of each atom from this plane.

Input (i) The number of atoms, ν ($\nu \leq 100$)

(ii) The atomic coordinates referred to orthogonal axes (output from Appendix 3.2.)

Output (i) Least-squares totals

(ii) Values of determinants used in solving the normal equations

(iii) Coefficients of the equation of the mean molecular plane

(iv) Perpendicular distances of atoms from the plane.

Time Approximately 2 minutes for 30 atoms.

System Alphacode Mark I

Method Let the equation of the mean molecular plane be

$$y = Ax + Bz + C \qquad \ldots\ldots\ldots\ldots (64)$$

ν observational equations may be set up which, by the standard least-squares procedure, reduce to 3 normal equations

$$\left.\begin{aligned} A\sum x^2 + B\sum xz + C\sum xy &= \sum xy \\ A\sum xz + B\sum z^2 + C\sum z &= \sum yz \\ A\sum x + B\sum z + \nu C &= \sum y \end{aligned}\right\} \qquad \ldots\ldots\ldots(65)$$

Solution of these equations is effected by evaluation of the four determinants

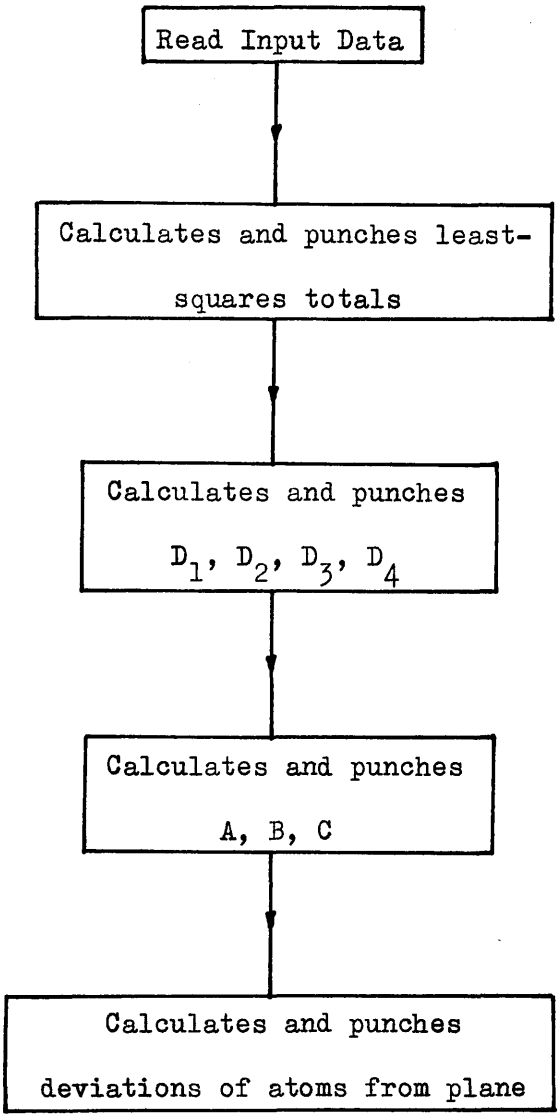
$$D_1 = \begin{vmatrix} \sum xy & \sum xz & \sum x \\ \sum yz & \sum z^2 & \sum z \\ \sum y & \sum z & \nu \end{vmatrix} \qquad D_2 = -\begin{vmatrix} \sum xy & \sum x^2 & \sum x \\ \sum yz & \sum xy & \sum z \\ \sum y & \sum x & \nu \end{vmatrix}$$

$$D_3 = \begin{vmatrix} \sum xy & \sum x^2 & \sum xz \\ \sum yz & \sum xz & \sum z^2 \\ \sum y & \sum x & \sum z \end{vmatrix} \qquad D_4 = \begin{vmatrix} \sum x^2 & \sum xz & \sum x \\ \sum xz & \sum z^2 & \sum z \\ \sum x & \sum z & \nu \end{vmatrix}$$

Then $A = D_1/D_4$, $B = D_2/D_4$ and $C = D_3/D_4$.

The perpendicular distance from a point (x_1, y_1, z_1) to the plane $y = Ax + Bz + C$ is given by $\dfrac{Ax_1 - y_1 + Bz_1 + C}{\sqrt{\{A^2 + B^2 + 1\}}}$

Flow Diagram



3.4. Programme for Conversion of Isotropic Temperature Factors to Pseudo-Anisotropic Temperature Factors.

Description The programme converts isotropic temperature factors to pseudo-anisotropic temperature factors for monoclinic systems.

Input (i) The number of atoms, ν.

(ii) ν isotropic temperature factors, B_θ.

(iii) a, b, c, $\cos\beta$, $\sin\beta$.

Output ν sets of pseudo-anisotropic temperature factors, B_{11}, B_{22}, B_{33}, B_{13}.

Time Approximately 2 minutes for 50 atoms.

System Alphacode Mark I.

Method For monoclinic systems let

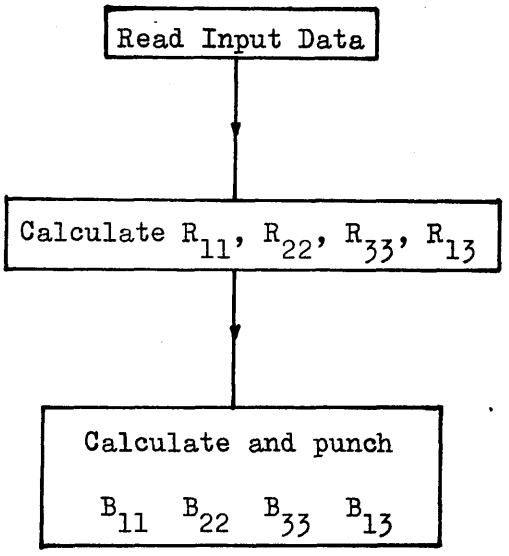
$$\exp(-B_\theta \sin^2\theta/\lambda^2) \equiv \exp\{-(B_{11}h^2+B_{22}k^2+B_{33}\ell^2+B_{13}h\ell)\} \quad ..(66)$$

Then $B_{11} = R_{11}B_\theta$, $B_{22} = R_{22}B_\theta$, $B_{33} = R_{33}B_\theta$, $B_{13} = R_{13}B_\theta$, ...(67)

where $R_{11} = \dfrac{1}{4a^2\sin^2\beta}$, $R_{22} = \dfrac{1}{4b^2}$, $R_{33} = \dfrac{1}{4c^2\sin^2\beta}$,

$$R_{13} = \frac{-\cos\beta}{2ac\sin^2\beta}$$

Flow Diagram



3.5. Programme for Calculation of Electron Density peak maxima

Description The programme calculates, according to the least-squares method described by Shoemaker et al (1950), the electron density peak maxima of a Fourier synthesis, assuming that the peak shapes can be fitted to a three-dimensional 10-parameter Gaussian function.

Input Positional matrix A in decimal form (see below)

Matrix h' in decimal form (see below)

Output Coordinate corrections to be applied to approximate peak maxima.

Time Approximately 5 minutes per atom.

System Sections 1, 3, 4 - Basic Standard Library Programmes

Section 2 - G.I.P.

Method The shape of the electron density peak is represented by the Gaussian function

$$\rho = \exp(p - \frac{r}{2}x^2 - \frac{s}{2}y^2 - \frac{t}{2}z^2 + ux + vy + wz + \ell yz + mxy + nxy) \quad \ldots\ldots\ldots\ldots(68)$$

$$\text{Thus } 2\log_e\rho = 2p - rx^2 - sy^2 - tz^2 + 2ux + 2vy + 2wz + 2\ell yz + 2mxz + 2nxy \quad \ldots\ldots\ldots\ldots(69)$$

27 points are chosen in a (3 x 3 x 3) parallelopiped as close as possible to the peak maximum and the central point of the 27 is taken as origin. Then, for each atom, using (69), 27 observational equations may be set up and can be represented in matrix notation by

$$A \xi = h \qquad \text{.....(70)}$$

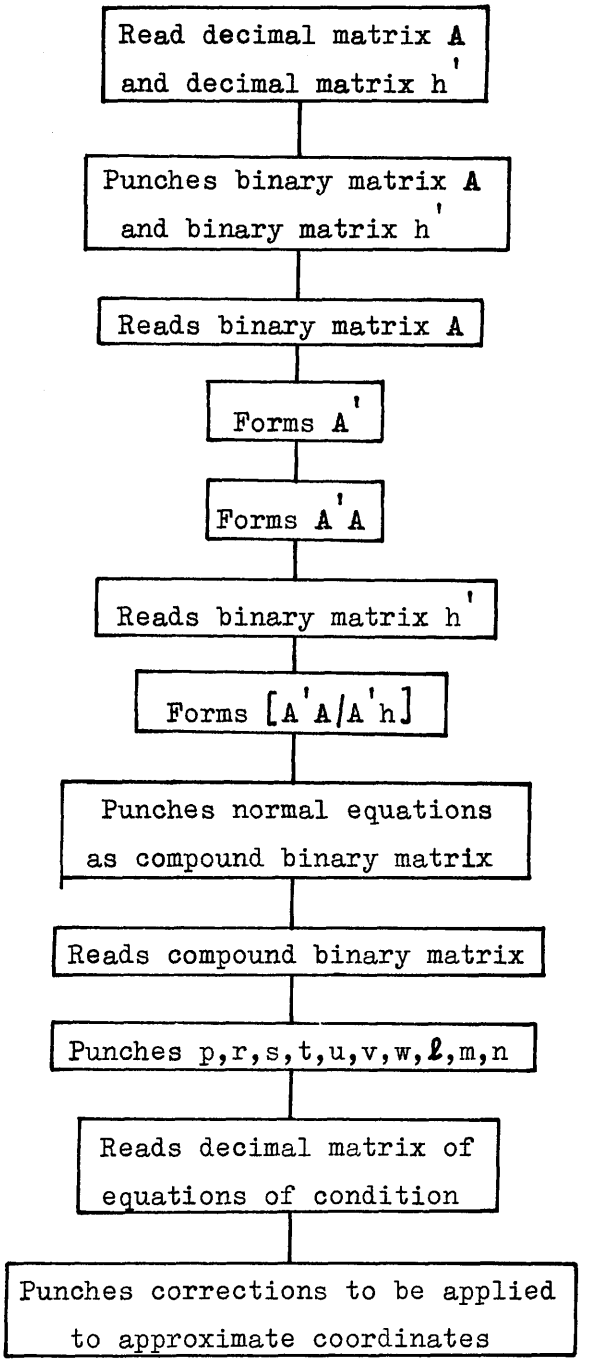
The corresponding normal equations may then be represented by

$$A' A \xi = A' h \qquad \text{.....(71)}$$

where A' is the transpose of **A**.

Solution of the normal equations yields the 10 parameters p, r, s, etc. The conditions for a peak maximum are that the partial derivatives of ρ (or better for computation, of $\log_e \rho$) with respect to x, y, z should be zero. Solution of the 3 equations of condition lead to the corrections which must be applied to the approximate coordinates chosen for the peak maximum.

Flow Diagram



3.6. Programme for the evaluation of the 'best' mean molecular plane through a set of atoms and the deviations of the atoms from this plane.

Description The programme calculates, by the method of least-squares, the equation of the 'best' plane through a set of point atoms subject to the condition that the weighted sum of the squares of the deviations, D_k, of points k from the plane be a minimum. The deviations are also evaluated by this programme.

Theory For the appropriate theory see Acta Cryst. (1959), 12, 600.

Practical Procedure

1. List the fractional coordinates x^1, x^2, x^3 and evaluate $\overline{x^1}$, $\overline{x^2}$, $\overline{x^3}$ according to the relationship

$$\overline{x^i} = \sum wx^i / \sum w ,$$

where w are the weighting factors.

2. Transform the atomic coordinates by the transformation $X^i = x^i - \overline{x^i}$.

3. Form the symmetric matrix A, where $a_{ij} = \sum X^i X^j$.

4. Form the matrix $\hat{A}$, the adjoint matrix of A.

5. Form the matrix $\hat{A}A = |A| I$. The diagonal elements of $\hat{A}A$ should be identical and equal to $|A|$. (The off-diagonal elements should be zero).

6. Evaluate the reciprocal unit cell translations $\vec{b}^{i}$

7. Form the symmetric matrix g, where $g^{ij} = \vec{b}^{i}.\vec{b}^{j}$

8. Form the matrix $B = \hat{A}g$.

9. Select the largest column vector of B and call it $m_{(0)}$.

10. Perform the iterative process $\quad Bm_{(0)} = m_{(1)}$

$$Bm_{(1)} = m_{(2)} \quad \text{etc.}$$

until the ratios of corresponding terms of $m_{(n)}$ and $m_{(n-1)}$ are constant. The final ratio will be referred to as the limiting ratio.

11. Calculate $\lambda^{(1)} = \dfrac{|A|}{\text{limiting ratio}}$

12. Assuming that $m_{(n)}$ represents the final column matrix, evaluate $gm_{(n)}$.

13. Evaluate the normalisation factor $\dfrac{1}{\sqrt{\tilde{m}_{(n)}\ gm_{(n)}}}$

where $\tilde{m}_{(n)}$ is the transpose of $m_{(n)}$.

14. Evaluate $m^{(n)}$ = normalisation factor x $m_{(n)}$.

15. Evaluate the origin-to-plane distance

$d = \tilde{m}^{(n)}\ \overline{x^{i}}$, where $\tilde{m}^{(n)}$ is the transpose of $m^{(n)}$.

16. Evaluate $D_k = \sum m_j^{(n)} x^j - d$, for each atom.

17. Evaluate $S = \sum_k w D_k^2$.

For strict coplanarity, $S = \lambda^{(1)} = 0$.

Input for Part I

(i) Number of atoms, ν $(\nu \leq 20)$.

(ii) ν sets of $^x/_a$, $^y/_b$, $^z/_c$.

(iii) ν values of w.

(iv) a, b, c, $\cos\beta^*$.

Output from Part I

(i) $\overline{x^i}$

(ii) Elements of **A**.

(iii) Elements of $\hat{\mathbf{A}}$.

(iv) Elements of $\hat{\mathbf{A}}\mathbf{A}$.

(v) $\overrightarrow{b^i}$.

(vi) Elements of *g*.

(vii) Elements of B.

Input for Part II

(i) Input data (i) - (iii) for Part I.

(ii) $\overline{x^i}$, $|A|$, *g*.

(iii) B, $m_{(0)}$.

Output from Part II

(i) $m_{(3)}$.

(ii) limiting ratios and average value.

(iii) $\lambda^{(1)}$

(iv) normalisation factor.

(v) origin-to-plane distance, d.

(vi) deviations, D_k.

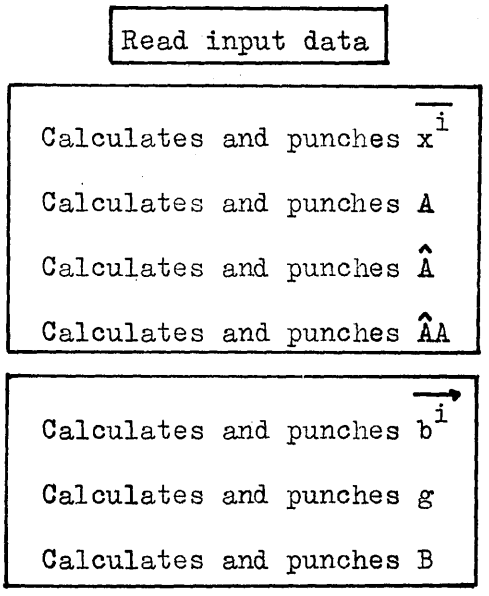
(vii) S.

Time Approximately 5 minutes for 20 atoms.

System Alphacode Mark I.

Flow Diagram

Part I



Part II

Reads Input data

Performs iterative process and punches $m_{(3)}$

Calculates and punches ratios at each stage and the limiting ratio

Calculates and punches $\lambda^{(1)}$

Calculates and punches $gm_{(3)}$

Calculates and punches the normalisation factor

Calculates and punches $m^{(3)}$

Calculates and punches d

Calculates and punches D_k

Calculates and punches S