Acta Cryst. (1953). **6**, 385

A Refinement of the Crystal Structure Analyses of Oxalic Acid Dihydrate

By F. R. Ahmed and D. W. J. Cruickshank

Chemistry Department, The University, Leeds 2, England

(*Received* 18 *June* 1952 *and in revised form* 14 *November* 1952)

Unsatisfactory agreement has previously been obtained for the dimensions of the oxalic acid dihydrate structure from the two sets of data of Robertson & Woodward, and of Brill, Hermann & Peters. By a new refinement, compatible results have been derived from the two sets of data. The methods used involved careful computational checks, corrections for the finite-series effect, a new method for determining atomic co-ordinates in unresolved projections, and statistical estimates of accuracy. The principal bond lengths given by the weighted mean results, with their estimated standard deviations, are C–C = 1·529±0·020 Å; C–O_1 = 1·285±0·012 Å; C–O_2 = 1·187±0·022 Å. The results, with other evidence, do not seem to support the hypothesis that the structure consists of oxonium and oxalate ions.

1. Introduction

1·1. *Historical survey*

Despite several attempts at modification, no very satisfactory agreement has so far been reached between the detailed dimensions of the structures of oxalic acid dihydrate derived from the two independent sets of experimental data of Robertson & Woodward (1936), (hereafter RW), and Brill, Hermann & Peters (1942), (hereafter BHP). The present paper is a contribution to the solution of this problem.

Following an early investigation by Wood (1926) and an approximate determination of the structure by Zachariasen (1934), RW measured 47 reflexions for the *x z* projection and 33 reflexions for the *y z* projection, using Cu radiation. They derived the structure I, Fig. 1, from experimental Fourier syntheses for these two projections. The *y* co-ordinates of the atoms C and O_2 were subject to some doubt, as these atoms were unresolved in the *y z* projection.

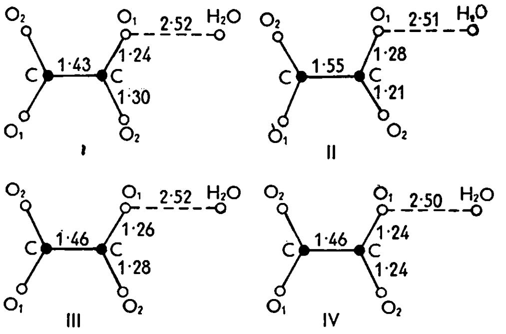

Fig. 1. Previous dimensions of oxalic acid dihydrate: (I) Robertson and Woodward. (II) Brill, Hermann & Peters. (III) Booth's modification. (IV) Dunitz & Robertson.

BHP, using Mo radiation, measured 212 reflexions for the *x z* projection, remeasured 16 of the *y z* previously measured by RW, and added 13 new planes. After making an allowance for finite-series effects, they derived the structure II, which shows a difference compared with I of 0·12 Å in the C–C bond and an inversion of the short and long C–O bonds.

Booth (1947) corrected the RW *x z* projection for finite-series errors by his back-correction method (Booth, 1945, 1946*a*), deriving the structure III.

Following this, Dunitz & Robertson (1947), by testing the structure-factor agreement on the two projections of a number of different structures, came to the conclusion that perhaps the most probable structure was of the form IV, in which the C–O bonds are very nearly equal.

During this work of Booth, and of Dunitz & Robertson, Finbak & Norman (1948, and private communication) developed a method for finite-series correction, somewhat akin to Booth's, which they applied to the *x z* projection.

In 1949 one of us (unpublished) tested the agreement between Booth's correction of RW's data and Finbak & Norman's correction of BHP's data and found that the difference in corrected co-ordinates was nearly twice what was to be expected from the likely errors of each.

1·2. *The present work*

The problem confronting us, then, was either to show that concordant results could be obtained from the two sets of data, or, if not, to find where the incompatibility arises. Our starting point has been the two sets of data; we have done no new experimental work (except a re-determination of the cell dimensions). The following features of our methods, though not all unique to this work, and the following results may be noted now:

(1) Where possible we have carried out parallel calculations with the two sets of data; in particular, for the *x z* projection we have compared results for the two sets of common reflexions.

(2) We have used computing techniques in which the errors of approximation should be negligible for the present purpose. In part we have found atomic positions by two essentially different computing techniques, so as to guard against some sources of mistake.

(3) A new method for dealing with unresolved peaks has been devised (Cruickshank, 1952) which we hope has been satisfactory for finding the y co-ordinates of the two unresolved peaks.

(4) We have made much use of statistical methods for estimating accuracy and for checking the compatibility of results derived either from different data or from different manipulative techniques.

(5) We conclude that concordant results can be obtained from the two sets of data, since the dimensions of the two structures we have derived agree within the estimated errors. From these we have derived a weighted mean structure whose accuracy approaches that obtained by a three-dimensional analysis with Cu data.

1·3. *Crystallographic data*

Oxalic acid dihydrate crystallizes in the monoclinic space group $P2_1/n$, with two molecules per unit cell of dimensions

$$a = 6{\cdot}119\pm0{\cdot}004,\ b = 3{\cdot}604\pm0{\cdot}003,$$
$$c = 12{\cdot}051\pm0{\cdot}005\ \text{Å},\ \beta = 106^\circ\ 16'\pm6'\ .$$

These dimensions have been obtained by us using the Straumanis method. They differ slightly from the dimensions $a=6{\cdot}12$, $b=3{\cdot}60$, $c=12{\cdot}03$ Å, $\beta=106^\circ\ 12'$ given by earlier workers. The C–C bond passes through the centre of symmetry.

Our calculations for the $x\,z$ projection were actually carried through in the more convenient set of axes given by $\mathbf{a}' = \frac{1}{2}(3\mathbf{a}+\mathbf{c})$, $\mathbf{c}' = \frac{1}{2}(-\mathbf{a}+\mathbf{c})$, which have $a' = 9{\cdot}465$, $c' = 7{\cdot}483$ Å, $\beta' = 91^\circ\ 42'$.

All the results given in this paper are quoted with reference to the original set of axes.

2. The x z projection

2·1. *Introduction*

For this projection the observed $|F|$'s, $|F_R|$ of RW and $|F_B|$ of BHP, were treated as independent sets of data. Two regions of reciprocal space were considered. Space (1) covered the observations common to RW and BHP, while space (2) covered all the BHP observations. Three sets of results were thus obtained,

RW (1): space (1), RW data;
BHP (1): space (1), BHP data;
BHP (2): space (2), BHP data.

Co-ordinates obtained from observed syntheses of RW (1) and BHP (1) will differ because of the different experimental errors of RW and BHP, and possibly because of slightly different apparent temperature factors. The difference between results from observed syntheses of BHP (1) and BHP (2) will be due to different finite-series errors and to the experimental errors having a larger effect on space (1) than on space (2).

2·2. *Scale and temperature factors*

For convenience, so as to need only one set of F_c for making the finite-series corrections to the two sets of F_o, the two sets of data were put on the same basis by multiplying F_R by a small temperature factor. We found

$$F'_R = \exp\,[0{\cdot}003s^2]\,.\,F_R$$

($s = 4\pi \sin\theta/\lambda$) to be the most suitable, no scale factor being necessary.

2·3 *Differential syntheses*

Observed differential syntheses (Booth, 1946*b*) were computed for RW (1), BHP (1) and BHP (2) using punched-card methods on Hollerith calculating machines. We found that the actual differences between the positions of the maxima of RW (1) and BHP (1) were compatible with the probable differences estimated statistically from a comparison of the two sets of F's. The details of a similar comparison based on our most accurate calculations are given in § 2·5. This substantially solved the problem of the compatibility of the results derived from the two sets of data. The differences between the results of either RW (1) and BHP (2), or BHP (1) and BHP (2) were larger than those between RW (1) and BHP (1), indicating the greater importance of finite-series and peak-overlapping errors.

Using Hartree f curves and the temperature factors obtained by Finbak & Norman (which we checked by least squares), structure factors F_{c1} were calculated with atomic positions close to the observed maxima of BHP (2). With these F_{c1}'s finite-series corrections for the two spaces were obtained by Booth's back-correction method from calculated differential syntheses. The corrections were much more important for space (1) than space (2), and on correction the discrepancies between the results of space (1) and space (2) were much reduced.

Actually in these differential syntheses we used, as the RW (1) data, the original F_R, and not the F'_R derived in § 2·2. Later results showed that the error of using one calculated model with the two sets of unadjusted data was not important. The F'_R were used in all later work.

2·4. *Fourier syntheses*

The results obtained from the differential syntheses were then checked, using ordinary Fourier syntheses computed at the points of a fine mesh; this involves an entirely different computing technique, and so should be a good check. As the positions of the observed and calculated maxima were known approximately, it was unnecessary to compute Fourier syntheses at 1/60ths or 1/120ths throughout the unit cell, but only to compute the density at the points of a fine mesh covering the areas of interest. Meshes of 4×4 points, with intervals of 3/1000ths of a cell side, were used. Errors of interpolation to the maxima should thus be small. The calculations were carried out by punched-card methods using four-figure trigonometric functions. The details of the methods used are being given by Ahmed (in preparation).

In all, 14 sets of calculations were made. We first computed observed syntheses for RW (1), BHP (1) and BHP (2), and calculated syntheses using the F_{c1} already obtained. The mean co-ordinate difference between these results and those from the differential syntheses was 0·002 Å, which we consider a satisfactory check. Two other observed syntheses for space (2) were also computed. In the first the RW data were taken for space (1), and the BHP data for the extra planes of space (2). In the second the means of the RW and BHP data were taken for space (1), and the BHP data for the extra planes. Both these syntheses gave positions very close to those of BHP (2).

A new set of structure factors F_{c2} was then computed, based on the corrected positions of this last mixed synthesis and using the Hartree f curves and the Finbak & Norman temperature factors. A set F^*_{c2} was also computed on these positions, using RW's composite f curve. RW's f curve is an experimental one in which carbon and oxygen have similar values in the ratio 6:10. This curve is only available for space (1), but gave rather better agreement between F_o and F_c. The plane (208) changed sign at this stage and was omitted from later calculations.

Further observed (without 208) and calculated syntheses with F_{c2} and F^*_{c2}, were then computed. The observed and corrected positions obtained from these syntheses are given in Table 1, and are taken as the final results for each set of data on this projection. The syntheses of F_{c2} and F^*_{c2} gave results agreeing within 0·001 Å, so that it was immaterial which was used.

2·5. *Accuracy and comparison of result*

In a well resolved projection the estimated standard deviation of an atomic co-ordinate, after finite-series correction, is (Cruickshank, 1949)

$$\sigma(x) = \frac{1}{A}\frac{2\pi}{a}\{\Sigma h^2(\Delta F)^2\}^{\frac{1}{2}}\Big/\left|\frac{\partial^2\varrho}{\partial x^2}\right|, \qquad (2\cdot1)$$

where A is the area of the cell in projection and $\partial^2\varrho/\partial x^2$ is the second derivative of the peak at its maximum. If we take $\Delta F = |F_o-F_c|$, we get an estimate of the combined experimental and residual finite-series errors (due to the imperfection of the calculated model). If we take $\Delta F = |F_B-F'_R|$, the formula gives an estimate of the expected difference

Table 1. *Oxalic acid dihydrate co-ordinates*

	Atom C			Atom O_1			Atom O_2			Atom $(H_2)O$		
	x/a	y/b	z/c	x/a	y/b	z/c	x/a	y/b	z/c	x/a	y/b	z/c
RW (1936)	−0·0367	0·0333	0·0508	0·0858	−0·0569	0·1478	−0·2280	0·2139	0·0350	−0·4520	−0·3917	0·1789
BHP (1942)	−0·0465	0·0615	0·0511	0·0828	−0·0560	0·1478	−0·2192	0·2435	0·0362	−0·4549	−0·3840	0·1780
xz projection; maxima of observed syntheses												
RW (1)	−0·0380	—	0·0488	0·0844	—	0·1466	−0·2245	—	0·0369	−0·4497	—	0·1789
BHP (1)	−0·0390	—	0·0500	0·0830	—	0·1464	−0·2243	—	0·0363	−0·4491	—	0·1787
BHP (2)	−0·0466	—	0·0511	0·0855	—	0·1482	−0·2201	—	0·0361	−0·4523	—	0·1796
xz projection; corrected co-ordinates												
RW (1)	−0·0441	—	0·0493	0·0867	—	0·1490	−0·2200	—	0·0371	−0·4544	—	0·1787
BHP (1)	−0·0451	—	0·0505	0·0853	—	0·1488	−0·2198	—	0·0364	−0·4537	—	0·1785
BHP (2)	−0·0458	—	0·0511	0·0850	—	0·1481	−0·2205	—	0·0363	−0·4524	—	0·1796
yz projection; corrected co-ordinates												
RW (1): Hartree f	—	0·0613	—	—	−0·0520	0·1490	—	0·2161	—	—	−0·3970	0·1805
RW (1): RW f	—	0·0440	—	—	−0·0528	0·1482	—	0·2163	—	—	−0·3960	0·1794
RW (1)—BHP (2)	—	0·0543	—	—	−0·0523	0·1488	—	0·2156	—	—	−0·3972	0·1799
Weighted mean co-ordinates												
RW (1)—BHP (2)	−0·0455	0·0543	0·0507	0·0854	−0·0523	0·1485	−0·2204	0·2156	0·0364	−0·4527	−0·3972	0·1796

between co-ordinates derived from the two sets of data for space (1).

In this projection the x- and z-co-ordinate errors are approximately equal. Table 2 gives the estimated errors of carbon and oxygen co-ordinates. The third line gives the root-mean-square differences of the carbon and oxygen co-ordinates found from the RW (1) and BHP (1) data by the second observed Fourier synthesis. The differences in the carbon co-ordinates are as large as the standard deviations estimated with $\Delta F = |F_B - F'_R|$, and the oxygen differences are rather smaller. Accordingly we conclude that the results from the RW and BHP space (1) data are consistent.

Table 2. *Accuracy data for x z projection*

	σ(C) (Å)	σ(O) (Å)
Space (1)		
$\Delta F = \|F'_R - F_c\|$	0·013 Å	0·010 Å
$\Delta F = \|F_B - F'_R\|$	0·012	0·010
Actual difference between RW (1) and BHP (1)	0·010	0·005
Space (2)		
$\Delta F = \|F_B - F_c\|$	0·0065	0·0055

The mean differences of the corrected RW (1) and BHP (2) co-ordinates are 0·016 Å for carbon and 0·010 Å for oxygen. Thus the two independent sets of data give results agreeing to within an amount to be expected from the standard deviations of each.

It can be seen from Table 2 that the BHP (2) co-ordinates are nearly twice as accurate as the RW (1) co-ordinates. Another aspect of the importance of the increased number of planes was shown by the synthesis mentioned in § 2·4, composed of the RW (1) data for space (1) together with the extra BHP planes, in which the co-ordinates differed only by a mean of 0·003 Å from the BHP (2) results.

The mean finite-series correction to each co-ordinate was 0·022 Å for carbon and 0·017 Å for oxygen in space (1), and 0·002 Å and 0·002 Å in space (2). Correspondingly, the discrepancies between BHP (1) and BHP (2) were reduced from 0·030 Å and 0·016 Å before correction to 0·006 Å and 0·006 Å afterwards. Thus the finite-series error for space (1), if not corrected, is more important than are the experimental errors.

3. The *y z* projection

3·1. *Introduction*

It is not convenient to deal with two independent sets of data for the $y\,z$ projection because BHP did not remeasure some of the planes measured by RW. We did, however, work with two sets of data. For space (1) we used the RW data, taking $F'_R = \exp[0{\cdot}003s^2]\,.\,F_R$. For space (2) we used all the data given by RW and BHP, taking a mean of F'_R and F_B for common planes.

The difficulty about this projection is that the atoms C and O_2 overlap. We found the y co-ordinates of these atoms by the modified differential Fourier method proposed by Cruickshank (1952), which invokes the criterion that the slopes of the observed and calculated electron densities should be equal at all the atomic positions. We took the z co-ordinates of C and O_2 as fixed, and as given by the $x\,z$ projection. The approximate equations of the modified differential Fourier method for the y co-ordinates of the unresolved atoms are a pair of simultaneous equations in the two co-ordinates; namely

$$\left.\begin{aligned}\left(\frac{\partial^2 K_{\mathrm{O}}}{\partial y^2}\right)_{\mathrm{O}}\varepsilon_{\mathrm{O}}+\left(\frac{\partial^2 K_{\mathrm{C}}}{\partial y^2}\right)_{\mathrm{O}}\varepsilon_{\mathrm{C}} &= -\left[\frac{\partial}{\partial y}(\varrho_o-\varrho_c)\right]_{\mathrm{O}},\\ \left(\frac{\partial^2 K_{\mathrm{O}}}{\partial y^2}\right)_{\mathrm{C}}\varepsilon_{\mathrm{O}}+\left(\frac{\partial^2 K_{\mathrm{C}}}{\partial y^2}\right)_{\mathrm{C}}\varepsilon_{\mathrm{C}} &= -\left[\frac{\partial}{\partial y}(\varrho_o-\varrho_c)\right]_{\mathrm{C}},\end{aligned}\right\}\quad(3{\cdot}1)$$

where ε_{O} and ε_{C} are refinements to trial values of the oxygen and carbon y co-ordinates, K_{O} and K_{C} are the calculated electron densities due to atoms O_2 and C, respectively, and their symmetry equivalents, ϱ_o and ϱ_c are the observed and calculated electron densities due to all the atoms, and the suffix $_{\mathrm{O}}$ or $_{\mathrm{C}}$ outside a bracket denotes that a derivative is to be evaluated at the trial position of O_2 or C.

3·2. *Fourier syntheses*

The computations for this projection were done by Fourier syntheses computed at points of a mesh covering the areas of interest, as in § 2·3. 4×4 meshes were used for the atoms O_1 and $(H_2)O$, but a mesh covering the whole region between C and O_2 was used for these unresolved atoms, with mesh intervals of 1/100th of the cell side.

y and z co-ordinates of O_1 and $(H_2)O$ were found from the maxima of observed syntheses, and were corrected for finite series by Booth's back-correction method from calculated syntheses. The terms on the right-hand side of the equations (3·1), required for the y co-ordinates of C and O_2, were found by numerical interpolation from observed and calculated syntheses; the coefficients on the left were found by direct computation.

Starting from an initial set of trial co-ordinates, the process described in the last paragraph was repeated twice, both for space (1) and space (2), using Hartree f curves. The co-ordinates derived from the third set of trial co-ordinates are shown in Table 1. The process was also carried through, from the third set of trial co-ordinates, for the space (1) data using the RW f curve; the results of this are also shown in Table 1.

3·3. *Accuracy and comparison of results*

The standard deviations of the y and z co-ordinates of the resolved O_1 and $(H_2)O$ atoms were estimated by equation (2·1). The estimation of the standard

Table 3. *Estimated standard deviations for y z projection*

(Values in Ångström units)

	C	O_1	O_1	O_2	$(H_2)O$	$(H_2)O$
	(y)	(y)	(z)	(y)	(y)	(z)
Space (1)						
$\Delta F = \lvert F'_R - F_{c3} \rvert$	0·049	0·021	0·011	0·030	0·020	0·013
$\Delta F = \lvert F'_R - F^*_{c3} \rvert$	0·053	0·023	0·010	0·033	0·022	0·012
Space (2)						
$\Delta F = \lvert F_{\text{mean}} - F_{c3} \rvert$	0·031	0·013	0·007	0·029	0·017	0·008

deviations of the y co-ordinates of C and O_2 is more involved. The solution of the equations (3·1) for ε_C is

$$\varepsilon_C = \frac{1}{D}\frac{2\pi}{bB}\Sigma\Sigma k(F_o - F_c) \times \left\{\left(\frac{\partial^2 K_O}{\partial y^2}\right)_O \sin\theta_C - \left(\frac{\partial^2 K_O}{\partial y^2}\right)_C \sin\theta_O\right\}, \quad (3·2)$$

where

$$D = \left(\frac{\partial^2 K_C}{\partial y^2}\right)_C \left(\frac{\partial^2 K_O}{\partial y^2}\right)_O - \left(\frac{\partial^2 K_C}{\partial y^2}\right)_O \left(\frac{\partial^2 K_O}{\partial y^2}\right)_C,$$

B is the area of the $y\,z$ projection, and $\theta_C = 2\pi(ky/b + lz/c)$ evaluated at the trial position of the carbon atom. This may be written

$$\varepsilon_C = \sum_0^K \sum_0^L a(kl)[F_o(0kl) - F_c(0kl)],$$

where the summation is only over symmetrically independent planes, and so, in this plane group, is a summation over positive values of the indices, and the $a(kl)$ are appropriate coefficients derived from (3·2). Accordingly, if the standard deviation of $F(0kl)$ is $\sigma(0kl)$, the standard deviation of ε_C, $\sigma(C)$, is given by

$$\sigma^2(C) = \sum_0^K \sum_0^L a^2(kl)\sigma^2(0kl).$$

By taking $\sigma(0kl) = |F_o(0kl) - F_c(0kl)|$ we obtain an estimate of the standard deviation of the unresolved carbon y co-ordinate. The standard deviation of the oxygen y co-ordinate was estimated similarly.

The estimated standard deviation of all co-ordinates obtained from this projection are shown in Table 3. F_{c3} and F^*_{c3} denote F's calculated from the third set of trial co-ordinates, using Hartree and RW f curves respectively.

Table 1 shows that the z co-ordinates of O_1 and $(H_2)O$ derived from the two projections are in satisfactory agreement both for space (1) and space (2), the mean difference between the two projections for space (2) being 0·008 Å.

There are considerable differences between the three determinations of the carbon y co-ordinate. The change from the Hartree to the RW f curve for space (1) shifts the position by 0·062 Å; however, the standard deviation of either determination is large, being 0·049 Å and 0·053 Å respectively. The space (2) determination gives an intermediate result of rather better accuracy.

4. Conclusions

4·1. *Weighted mean results*

If several independent estimates of a co-ordinate are available, each having a standard deviation σ_i, the weighted mean estimate of the co-ordinate is

$$\bar{x} = (\sum_i w_i x_i)/\sum_i w_i, \quad (4·1)$$

where $w_i = 1/\sigma_i^2(x)$. The standard deviation σ_m of the weighted mean is then given by

$$1/\sigma = \sum_i 1/\sigma_i^2. \quad (4·2)$$

For the RW (1) data, the final z co-ordinates for O_1 and $(H_2)O$ are taken to be the weighted means of the estimates from the two projections. (Strictly this procedure is not quite accurate as the z co-ordinates derived from the two projections are partially correlated through the $(00l)$ planes, which are common to both projections; however, the error of this procedure will be negligible.) The RW (1) data gives only a single estimate of all other co-ordinates. Final results are obtained from the BHP (2) data in a similar way.

Since the RW and BHP data do not lead to inconsistent results, a final weighted mean set of co-ordinates can be obtained by taking both together. The final weighted mean x co-ordinates are taken to be the weighted mean of the RW (1) and BHP (2) results from the $x\,z$ projection, weighted according to the standard deviations of Table 2. This assumes that these results are independent. The observed F's are certainly so, but not the finite-series corrections, since a common model is used for both spaces. However, since the corrections are very different for the two spaces, we think that no serious mistake will be made by this method. The final y co-ordinates are taken as those obtained from space (2), with the RW−BHP data, on the $y\,z$ projection.

The final weighted mean of the z co-ordinates of O_1 and $(H_2)O$ are taken as the weighted mean of the RW (1) and BHP (2) results from the $x\,z$ projection and the RW−BHP (2) results from the $y\,z$ projection. The final weighted mean of the z co-ordinates of C and O_2 are taken as the weighted mean of the RW (1) and BHP (2) results from the $x\,z$ projection.

The estimated standard deviations of the final weighted mean co-ordinates are given in Table 4.

Table 4. *Estimated standard deviations of the final weighted mean co-ordinates*

(Values in Ångström units)

	$\sigma(x)$	$\sigma(y)$	$\sigma(z)$
C	0·0061	0·0308	0·0052
O_1	0·0051	0·0130	0·0042
O_2	0·0052	0·0295	0·0046
$(H_2)O$	0·0052	0·0170	0·0040

Tables 5 and 6 give the bond lengths and angles derived from the final corrected RW, BHP, and weighted mean results. For comparison with values given by RW (1936) and BHP (1942), these results are given using the old axial lengths. The final weighted mean results are also given using the new axial lengths.

If the positions of the two atoms forming a bond are independent the standard deviation of the bond length is given by

$$\sigma^2(l)=\{\sigma^2(x_1)+\sigma^2(x_2)\}\cos^2\alpha+\{\sigma^2(y_1)+\sigma^2(y_2)\}\cos^2\beta + \{\sigma^2(z_1)+\sigma^2(z_2)\}\cos^2\gamma\ , \qquad (4{\cdot}3)$$

where $\sigma(x_1)$, $\sigma(y_1)$, $\sigma(z_1)$ are the standard deviations of the co-ordinates of the first atom and cos α, cos β and cos γ are the direction cosines of the line joining the atoms. For the bond C–O_2, the y co-ordinates of the atoms are not independent, and in equation (4·1) $\{\sigma^2(y_1)+\sigma^2(y_2)\}$ is replaced by $\sigma^2(y_C-y_{O_2})$, which is computed directly after the manner of § 3·3.

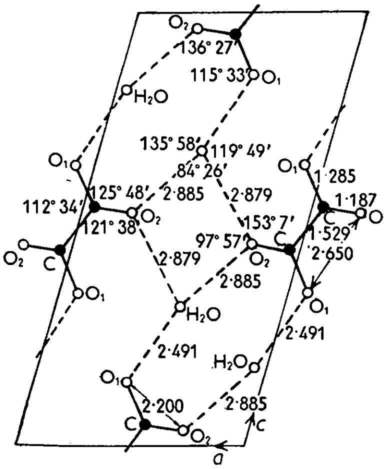

Fig. 2. Dimensions of the oxalic acid dihydrate structure.

If the positions of the three atoms 1, 2, 3 forming an angle at 2 are independent, the standard deviation of the angle is given by

Table 5. *Bond lengths and estimated standard deviations*

(Values in Ångström units)

	C–C	C–O_1	C–O_2	$(H_2)O$–O_1	$(H_2)O$–O_2	$(H_2)O$–O_2
RW (1936)	1·43	1·24	1·30	2·52	2·84	2·87
BHP (1942)	1·55	1·28	1·21	2·51	2·83	2·86
Present results						
RW corrected	1·468	1·287	1·216	2·491	2·878	2·874
BHP corrected	1·537	1·279	1·186	2·486	2·881	2·878
Final weighted mean	1·527	1·284	1·187	2·487	2·881	2·878
Final weighted mean (using the new axial lengths)	1·529	1·285	1·187	2·491	2·885	2·879
Estimated standard deviations						
RW corrected	0·038	0·021	0·041	0·016	0·025	0·016
BHP corrected	0·021	0·012	0·023	0·009	0·018	0·010
Final weighted mean	0·020	0·012	0·022	0·008	0·018	0·009

Table 6. *Angles*

	C–C–O_1	C–C–O_2	O_1–C–O_2	Sum of angles
RW (1936)	116°	120° 30′	123° 30′	360°
Present results				
RW (1) corrected	114° 42′	121° 41′	123° 26′	359° 49′
BHP (2) corrected	112° 23′	121° 17′	126° 20′	360° 0′
Final weighted mean	112° 28′	121° 39′	125° 53′	360° 0′
Final weighted mean (using the new axial lengths)	112° 34′	121° 38′	125° 48′	360° 0′
Estimated standard deviations				
RW (1) corrected	2° 5′	2° 5′	1° 23′	
BHP (2) corrected	1° 14′	1° 5′	0° 51′	
Final weighted mean	1° 13′	1° 4′	0° 49′	

$$\sigma^2(\theta) = (1/lm \sin\theta)^2[(x_2-x_3)^2\sigma^2(x_1)+(x_1-2x_2+x_3)^2\sigma^2(x_2) + (x_2-x_1)^2\sigma^2(x_3) + \text{similar terms in } y \text{ and } z]\,, \quad (4\cdot4)$$

where l and m are the lengths of the two bonds, θ is the angle, and x, y, and z are co-ordinates referred to orthogonal axes. For angles including C and O_2, the correlation of the y co-ordinates of these atoms was taken into account. The estimated standard deviations of the bond lengths and angles are given in Tables 5 and 6.

Fig. 2 shows the distances and angles for the molecule and its neighbours, as computed from the final weighted mean co-ordinates.

The percentage disagreements between observed structure factors and those calculated for the final weighted mean positions were as follows:

$x\,z$ projection

Space (1): RW data (F_R), 14·6 using Hartree f's, 10·9 using RW f's.
BHP data, 11·7 using Hartree f's, 10·9 using RW f's.
Space (2): BHP data, 14·4.

$y\,z$ projection

Space (1): RW data, 14·5 using Hartree f's, 12·1 using RW f's.
Space (2): RW−BHP data, 13·9.

On the $x\,z$ projection the RW data gave 10·4 using RW f's and the corrected RW (1) positions.

4·2. *Discussion of methods*

It seems from this work that no great error is introduced by neglecting to correct the positions of atoms in resolved projections for termination of series when data from Mo radiation are used. The correction is important when only data from Cu radiation are used. With unresolved projections large errors can be made, with either radiation, if the Fourier maps are not interpreted correctly. The final results we have obtained are generally much closer to those given by BHP (1942) than to those originally given by RW (1936). There is no reasonable doubt that the bond C–O_1 is longer than C–O_2; when the various corrections described in this paper have been made, this result is also indicated by the RW data alone, though the estimated errors are then too large to guarantee the conclusion.

A final accuracy approaching that of three-dimensional work with Cu radiation has been given by the two-dimensional projections using Mo data; but the accuracy of any particular bond depends considerably on its inclination to the unresolved projection.

The experience obtained from this work seems to indicate that many structures already solved by two-dimensional methods may not be as accurate as has been supposed; but that, without any further experimental work, important improvements could be made by correcting for finite series and by systematic solution of unresolved projections. However, even then, the resultant accuracy may be only one-half or one-third that of three-dimensional work.

4·3. *Discussion of results*

The dimensions of the structure as given by the final weighted mean results are of considerable interest. Particularly relevant is a comparison with the results of the recent re-investigation of the structure of α-anhydrous oxalic acid by Cox, Dougill & Jeffrey (1952). Their results, with the estimated standard deviations, are given in Table 7; the notation there has been chosen to correspond to that of this paper, and interchanges O_1 and O_2 as given by Cox *et al.* The general dimensions of the molecule in both structures are very similar; in both cases, the central C–C bond does not differ significantly from the standard C–C single-bond length of 1·5445 Å in diamond; in both cases, there is a short C–O bond of about 1·19 Å, and a longer one of about 1·29 Å, the short bond being associated with a C–C–O angle of about 122°, and the longer C–O bond with an angle of about 111°.

Table 7. *Cox, Dougill & Jeffrey's results for α-anhydrous oxalic acid*

Bond	Length	Standard deviation
C–C	1·560 Å	0·0103 Å
C–O_1	1·289	0·0068
C–O_2	1·194	0·0067

Angle	Value	Standard deviation
C–C–O_1	109° 13′	0° 33′
C–C–O_2	122 32	0 36
O_1–C–O_2	128 9	0 31

On detailed comparison the only one of these five independent parameters which differs importantly in the two structures is the angle C–C–O_1; the dihydrate has 112° 34′ with standard deviation 1° 13′, and the α-anhydrous 109° 13′ with standard deviation 33′. The ratio of this difference to the standard deviation of the difference is (3° 21′)/(1° 20′) = 2·50. The chance, on the hypothesis that the angles are equal, that a difference as large or larger than this could arise owing to random errors is $1-\text{erf}\,(2\cdot50/\sqrt{2}) = 0\cdot0125$. The difference is accordingly possibly significant. However, the general similarity of the dimensions of the oxalic acid molecule in the two structures remains, the small differences being possibly due to the different hydrogen bonding systems. This similarity is strong evidence against the hypothesis that the dihydrate should be regarded as built up of $(H_3O)^+$ and oxalate ions.

This result is also confirmed by the work of the Pasadena school on amino acids. These are the only other analyses of structures involving carboxyl groups known to us in which care has been given to the correction of finite-series errors, etc. Table 8 gives the dimensions of the carboxyl groups in alanine (Donohue, 1950), serine (Shoemaker, Barieau, Donohue & Lu,

Table 8. *Dimension of carboxyl group in aminoacids*

	C–C–O	Long C–O	Short C–O	C–C–O
Alanine	113°	1·27 Å	1·21 Å	121°
Serine	117	1·27	1·26	117
Threonine	116	1·25	1·24	117
α-Glycylglycine	115	1·26	1·22	117
β-Glycylglycine	112	1·27	1·21	123
Hydroxyproline	115	1·27	1·25	119
N-Acetyglycine	112	1·31	1·19	124

1953), threonine (Shoemaker *et al.*, 1950), α-glycylglycine (Biswas, Hughes & Wilson, private communication), β-glycylglycine (Hughes & Moore, 1949), hydroxyproline (Donohue & Trueblood, 1952) and N-acetylglycine (Carpenter & Donohue, 1950). All this work is of high accuracy, the standard deviations of the bond lengths in serine and threonine being of the order of 0·010 Å. The first six amino acids listed in Table 8 are all internal salts (*Zwitterions*), as is established by the disposition of hydrogen bonds, and these have the two C–O bond lengths much more nearly equal than in N-acetylglycine. Part of the variation among these six is to be attributed to the particular intermolecular attractions. The dimensions of the carboxyl group in N-acetylglycine are very similar to those in the α-anhydrous and dihydrate oxalic acid structures. It is also relevant that the longer C–O bond of 1·31 Å in N-acetylglycine is associated with an oxygen–oxygen hydrogen bond of 2·56 Å, and that the longer C–O bond of 1·285 Å in the dihydrate is associated with a hydrogen bond of 2·491 Å.

On this evidence of the similarity of the dimensions of the carboxyl group in the dihydrate with those in N-acetylglycine, rather than with those in the *Zwitterion* amino acids, we again infer against the dioxonium oxalate hypothesis.

Arguments based solely on bond lengths are necessarily only an indirect discussion of the hypothesis. We endeavoured to locate the hydrogen atoms by the use of a $(\varrho_o-\varrho_c)$ difference map, hoping to find to which oxygen atoms the hydrogens were closest. The interpretation of the map was too uncertain for any significant conclusions. Nor, having estimated the probable errors in the electron density, do we think that any significant conclusion can be drawn, from the difference map given by Finbak & Norman (Fig. 7), as to whether in the short 2·491 Å hydrogen bond, the hydrogen is closer to the carboxyl oxygen or to the water oxygen.

The indirect conclusion from X-ray structure analysis agrees with the conclusion reached by Richards & Smith (1951) from an examination of the nuclear magnetic resonance spectrum. The proton resonance absorption curve is compatible with a structure consisting of water molecules and oxalic acid molecules, and cannot be fitted with a structure involving oxonium ions. Marignan (1948) has interpreted the Raman spectra of the solid on the basis of a structure containing oxalic acid and water molecules.

The observations that the central C–C bonds in the α-anhydrous and dihydrate oxalic acid structures are close to the standard C–C single bond length suggests that there is no appreciable degree of conjugation across this bond, and that some other explanation must be sought for the planarity of the oxalic acid molecule. This question has been discussed recently by Jeffrey & Parry (1952).

Our thanks are due to Prof. E. G. Cox for his interest and encouragement and for the computing facilities available in this laboratory, to Miss M. W. Dougill for her assistance with the Straumanis measurements, and to Dr N. Norman for providing us with some of his results.

References

Booth, A. D. (1945). *Nature, Lond.* **156**, 51.

Booth, A. D. (1946*a*). *Proc. Roy. Soc.* A, **188**, 77.

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

Booth, A. D. (1947). *Proc. Roy. Soc.* A, **190**, 490.

Brill, R., Hermann, C. & Peters, A. (1942). *Ann. Phys. Lpz.* **42**, 357.

Carpenter, G. B. & Donohue, J. (1950). *J. Amer. Chem. Soc.* **72**, 2315.

Cox, E. G., Dougill, M. W. & Jeffrey, G. A. (1952). *J. Chem. Soc.* p. 4854.

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

Cruickshank, D. W. J. (1952). *Acta Cryst.* **5**, 511.

Donohue, J. (1950). *J. Amer. Chem. Soc.* **72**, 949.

Donohue, J. & Trueblood, K. N. (1952). *Acta Cryst.* **5**, 419.

Dunitz, J. D. & Robertson, J. M. (1947). *J. Chem. Soc.* p. 142.

Finbak, C. & Norman, N. (1948). *Acta chem. scand.* **2**, 813.

Hughes, E. W. & Moore, W. J. (1949). *J. Amer. Chem. Soc.* **71**, 2618.

Jeffrey, G. A. & Parry, G. S. (1952). *Nature, Lond.*, **169**, 1105.

Marignan, R. (1948). *Bull. Soc. chim. Fr.* p. 352.

Richards, R. E. & Smith, J. A. S. (1951). *Trans. Faraday Soc.* **47**, 1261.

Robertson, J. M. & Woodward, I. (1936). *J. Chem. Soc.* p. 1817.

Shoemaker, D. P., Barieau, R. E., Donohue, J. & Lu, C-S. (1953). *Acta Cryst.* **6**, 241.

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

Wood, J. F. (1926). *Proc. Univ. Durham Phil. Soc.* **7**, 111.

Zachariasen, W. H. (1934). *Z. Krystallogr.* **89**, 442.