

MIYAKE: They are to be found in the paper by Kohra & Kikuta (1968). Their further studies will be published shortly.

JENNINGS: The agreement between *measured* and *calculated* F -values which you have shown is, in several cases, rather better than the level of agreement between different sets of *measured* F -values evident in the Powder Project results which I showed yesterday (Paper G2·1). I would like to

know whether in the cases you presented the B -factors were the same in all experiments or whether they were adjusted to give the best agreement of F_o and F_c .

MIYAKE: Different B -values were used in different experiments but this does not necessarily imply that artificial adjustments were made to give the best agreement of F_o and F_c for all reflexions.

Acta Cryst. (1969). A25, 264

H1·3

The Role of Intensity Measurement Projects

BY A. MCL. MATHIESON

Division of Chemical Physics, C.S.I.R.O., Chemical Research Laboratories, Melbourne, Australia

The accuracy of the determination of X-ray intensities, and hence structure factor F -values, is of crucial importance to studies of the solid state, particularly at the present time when computational facilities allow exhaustive analysis of experimental data against theoretical models. Experimental errors are far less easy to estimate than appears from consideration of published individual experiments at a first, or even more careful, consideration; methods of both estimating and correcting for them need very careful elucidation. By far the best way to do this – and probably the only way that will yield reliable information as to where we are in this ill-defined field – is to organize group projects specifically to allow estimation of the magnitude of overall error, the identification of individual sources of error where possible and hence, the detection of specific experimental features, which should be either carefully assessed in each experiment to allow the magnitudes of errors to be kept as low as possible, or actually physically corrected. It is helpful in this regard to allow error-sources to be thought of as stabilized or variable according to the type of project planned. It is obvious that careful planning of a series of projects with different characteristics will be an even more powerful tool for the investigation of error-sources in depth. In this paper, two projects are considered in some detail – one originated by the I.U.Cr., the other by the A.C.A. – both having considerable similarities but also instructive differences. A method of comparing and also of editing the projects is considered using the correlation R factor $|R_{ij}|$ where $|R_{ij}| = \frac{\sum |F_i - F_j|}{\frac{1}{2} \sum |F_i + F_j|}$. This is seen to allow simple but important deductions.

Our basic concern is with the individual experiment from which structure factor F -values are derived – how to estimate its accuracy, to detect the error-sources responsible for its failure to attain greater accuracy and, having learned about these, to devise means to achieve the necessary improvement.

Most factors capable of producing errors in the measurement of intensities and hence F -values, have probably been recognized. However, the assessment of the magnitude of their individual contributions to errors in measurement is not, in each case, necessarily well-defined. In addition, there is the possibility of other factors whose effects may have been underestimated and hence have not been fully explored. The elucidation of the influence of a wide range of factors, in a piecemeal fashion, by individual experiments, although obviously of great value, is extremely tedious. Furthermore such an approach to the task of determining the overall error magnitude does pose certain problems as, I trust, will become evident.

In particular, each scientist tends to be optimistic regarding the accuracy of his own measurements in that he feels that he has adequate knowledge and con-

trol of the variables involved in his experiments. Moreover since few experiments are repeated, the normal single experiment is unique and its accuracy is largely indeterminate although an estimate of precision from the internal details of the experiment can be made; but this may be optimistic if the full range of error-sources in the experiment is not appreciated. It is perhaps unfortunate that for such isolated situations, F -values calculated from atomic models have been used as a guide to the assessment of accuracy of experimental measurements. If the models are inadequate, the attempted estimate made in this manner can be misleading.

Opinions on levels of accuracy have been expressed on earlier occasions but such estimates remain personal assessments without backing numerical evidence.

A counterbalance to individual bias and a somewhat less subjective assessment of accuracy can be attained by comparison of F -values given in the literature. Unfortunately, opportunities to apply this procedure are few since it is only rarely that experimenters measure and publish new values concerning compounds already reported, except to prove some point in the de-

velopment of a theoretical theme. The situation, concerning *experimental* values, in such cases is often further confused by their being assessed against the theoretical calculations. Recently, Maslen (1967) has shown the possibilities inherent in a careful statistical analysis of the published experimental data on NaCl. He was able to show that the sets of measurements could be differentiated into a major consistent sub-group and a minor non-consistent sub-group. On the basis of this analysis, he pointed to trends and possible reasons for the deviations in individual sets of the latter group.

The possibility of using the literature as a source of material for analysis is limited. Sample sizes are bound to be small and, if a number of data-sets are available, they probably originate from different techniques. It would be more valuable to study the situation under controlled conditions. An organized arrangement to perform a number of experiments of closely similar type – a project – could yield useful information on the magnitude and perhaps types of factor which militate against the derivation of structure factors of high accuracy.

Error-sources and project types

It is necessary to point out that, of the error-sources likely to exist in project studies, some can be labelled *variable* and some *stabilized*, their relative distribution varying with the type of project. To clarify the use of these terms, let us take an example. One project involves measurement, by n different persons, of data from a single crystal of a material X , while a second project involves measurement by n different persons each on his own single crystal, also of the same material. In the first project, all are measuring the same crystal so potential major error-sources due solely to the crystal (such as absorption, extinction) are the same for all participants and such error-sources are what I refer to as *stabilized*. In the second project, the error-sources due to the crystal are *variable*. This viewpoint can be extended to deal with other error-sources. In this respect, an *individual experiment* can be considered as a complete collection of *stabilized error-sources*. This viewpoint focuses attention on the dangers of relying too much on the results of an individual experiment.

Having established the distinction between *variable* and *stabilized* error-sources, it is therefore evident that projects can be classified according to how many error-sources involved are variable and how many stabilized. For our present purposes, let us arbitrarily delineate a possible group of projects, Table 1. I have not designated an extensive list of error-sources but have treated these in groups or regions, such as (a) the diffractometer, (b) the crystal and (c) features involving the interaction between them. To each, a measure of the overall magnitude of regional errors is allocated *e.g.* r_a , in respect of the diffractometer. The subdivision is arbitrary but still allows us to indicate the extremes, namely the individual experiment with no variable error-source, in the sense used here, and, at the other end, a project where the maximum number of variable error-sources is involved.

A project of type 1, Table 1, involves measurement, on one diffractometer assembly, by one operator or group, of data for n different crystals of one compound. A type 1 project focuses attention on the crystal as variable error-source. No such formal project involving inter-comparison of *experimental F-values* has been carried out but the results of Abrahams (1964) and Abrahams & Bernstein (1965) on 5 crystals of NaCl contain the essential information and invite re-consideration from this viewpoint.

A project of type 2 involves measurement by n persons of 1 crystal. This corresponds to the main component of the American Crystallographic Association (A.C.A.) project with $n=7$, a report on which has been published by Abrahams, Alexander, Furnas, Hamilton, Ladell, Okaya, Young & Zalkin (1967).

Project type 3 involves measurement by n persons, each using a different crystal. This corresponds to the International Union of Crystallography (I.U.Cr.) project, with $n=16$. It should also be pointed out that the second component of the A.C.A. project yielded a limited range of data of this type – comparisons of 5 secondary crystals with the primary crystal being reported.

We are only slowly learning to appreciate the relationship between projects. Ideally, a hierarchy of projects, Table 1, should be carried out on crystals of one compound. In the case of the actual projects undertaken, one was on a highly symmetrical inorganic

Table 1

Total of regional errors	Diffractometer (a)	Diffractometer—crystal interaction (c)	Crystal (b)	Measure of sum total of variables
Individual experiment	1	?	1	
Project 1	1	?	n	$ R_{ij} (b+c)$
Project 2 (\equiv A.C.A.)	n	?	1	$ R_{ij} (a+c)$
Project 3 (\equiv I.U.Cr.)	n	?	n	$ R_{ij} (a+b+c)$
Peak $ R_{ij} (a+b+c) > R_{ij} (a+c)$				
Peak $ R_{ij} (a+b+c) > R_{ij} (b+c)$				

compound, CaF_2 , and the other on a low-symmetry low absorption organic compound, D(+)-tartaric acid. The experience in both projects is that the selection of a standard crystal which is stable, chemically and physically, has suitable external morphological features and also internal morphology (if I may use that phrase to cover extinction characteristics) and is not damaged by X-radiation has proved a difficult proposition.

Basic features of projects

A project has to arrange for (a) the collection of data, (b) their reduction to a comparable basis and analysis and (c) the derivation of conclusions, or perhaps more cautiously, inferences.* There appear to be two principal reasons for the initiation of projects involving the measurement of intensities and their reduction to F -values.

(i) An assessment of the *spread of F -values* derived as a result of application of a particular technique or group of techniques *i.e.* to gain a measure of the consistency of these measurements not involving gross atypical error-sources. For this purpose, it is likely that procedures involving scaling between data-sets may be important.

(ii) Location by analysis procedures of (a) which regions require detailed study and improvement and (b), if possible, which aspects of technique require to be rectified. We are here interested in trends and cross-correlations and scaling is of lesser significance. Concerning analysis of projects, we need to recognize that a completely objective and final assessment of a project is unattainable. However esoteric the statistical analytical procedure applied, the conclusions depend on the questions asked, the appropriateness of the design of the project for answering these questions and especially on the analysts' appreciation of the design and operational details of the many apparatus components involved in the measurements.† In general, my approach is simply that a project analysis should be useful and that this is reflected in the significance of the inferences reached. These inferences, adequately supported by the analysis evidence, should be instructive about the current state of accuracy and/or point to definite aspects of technique which require investigation *i.e.* a project should be didactic and clearly give guidance for future action (*vide* Schomaker, 1965).

* The subject of the design of a project in respect of what data and associated information is requested under (a) above is a large one and I will not attempt to treat it here (*cf.* Fisher, 1947).

† Given an array of project data, basically there can be as many analyses as there are individuals willing to be analysts and they may well all arrive at different conclusions. There is nothing sacrosanct about an analysis or a report. In fact, the more people taking a close interest in the project data, the greater will be both the stimulus of such projects and their contribution to the clarification of the many problems involved in reducing error.

As indicated, there are many procedures of analysis which can be applied to project data. No one procedure can reveal all the points we may wish to bring to notice. Rather than attempt the task of summarizing the various analysis procedures available, I want to give some coherent picture of the relationships, in the round, between projects and also indicate the relationships between experimental measurements, derived at different project levels, and absolute values. I intend to use a device which is suited to my purpose. In its present form, it has not been elaborated but perhaps its simplicity may bring home the relationships between values derived under the conditions outlined in Table 1.

The $|R_{ij}|$ frequency distribution

The procedure used is designed to lay stress upon the differences between pairs of experiments, i and j , in a project.* The index used is $|R_{ij}|$ as given in (1)†:

$$|R_{ij}| = \frac{\sum (F_i - F_j)}{\frac{1}{2} \sum (F_i + F_j)}. \quad (1)$$

There is an obvious similarity to the conventional index in structural crystallography, $R = \sum |F_o - F_c| / \sum F_o$. This index $|R_{ij}|$ provides a *single-valued overall assessment of 'distance'* between each pair of sets of experimental data. The frequency of occurrence of values of $|R_{ij}|$ plotted against R , the frequency distribution, provides a means of assessment not only of the project as a whole but can be instructive concerning individual sets. Its effectiveness tends to increase as n increases.

It is useful to outline the significance of the $|R_{ij}|$ frequency distribution for ideal situations first and progress to real situations.

(a) Let us imagine a project involving n diffractometers, each being an ideal instrument, perfectly handled. Whether we are dealing with a type 2 or 3 project is immaterial at this stage. With all conditions ideal, each set of F -values will be identical. The resultant array of $|R_{ij}|$ values will be identically zero so the frequency distribution will be a delta function at $R=0$, Fig. 1(a).

(b) Imagine these ideal experimenters with instruments which have become somewhat worn with usage so that they all have marginally small errors – contributed from the various different sources in different proportions in each experiment. The resultant frequency distribution reflects this change in that the delta

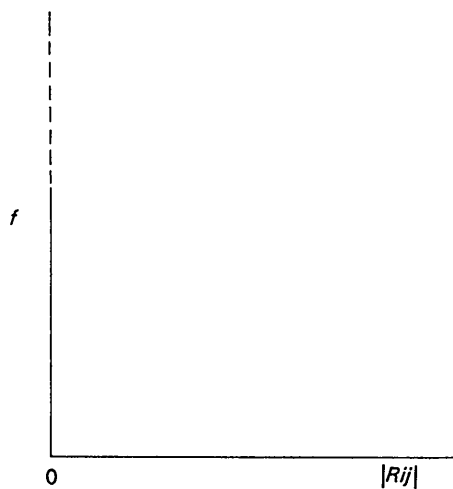
* Other procedures make comparisons between the data in set i with the mean values, μ , derived from all sets. Pair analysis is, for our purposes, rather more searching in respect of outliers. Deviations from the mean are perhaps more appropriate in dealing with homogeneous sets. Comparison with a mean set yields only n indices, $|R_{i\mu}|$, while the pair procedure yields $n(n-1)/2$ indices $|R_{ij}|$.

† It is recognized that (1) could be replaced by a relationship of the form such as (1a), which would be more acceptable to conventional statistics.

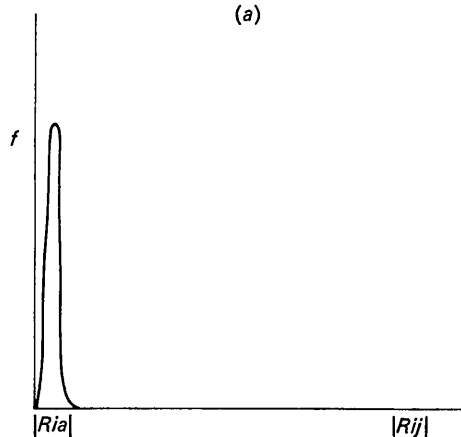
$$2|R_{ij}| = \frac{\sum |F_i - F_j|^2}{\frac{1}{2} \sum (F_i^2 + F_j^2)}. \quad (1a)$$

However, (1) is retained because of its similarity to the traditional index, R , and its adequacy to the present purpose.

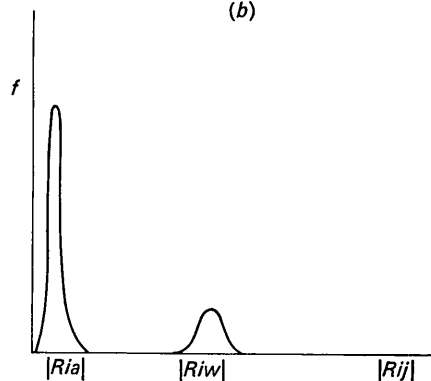
function is now replaced by a distribution (shape not defined) whose peak will be at some R value, R_1 , Fig. 1(b), which can be regarded as a function of the mean



(a)



(b)



(c)

magnitude, r_1 , of the distribution of error contributions.*

(c) Let us imagine that in this company of n crystallographers, one (say) has become weary of ensuring all operational details are just so and has been less careful in one or more of the basic adjustments. As a result, all indices R_{tw} involving his set w will be larger and the effect will be evident in the array and in the distribution, Fig. 1(c). From the distribution, the separation into two groups is evident – (a) the main group (excluding w) with a peak value of R_{1a} and a minor set (in this case, one) at R_{tw} due to set w . This situation brings us face to face with one of the fundamental questions in analyses. Is the main group correct and the outlier wrong or is it the other way round? I will return to this matter in a moment. Removal of set w from the project will cause reversion to the distribution in Fig. 1(b) and also the value, R_{1a} which may have been shifted by the change in scale consequent on the incorporation of set w , reverts to R_1 .

* Consideration of the possible functional relationship between r_1 and R_1 need not concern us here.

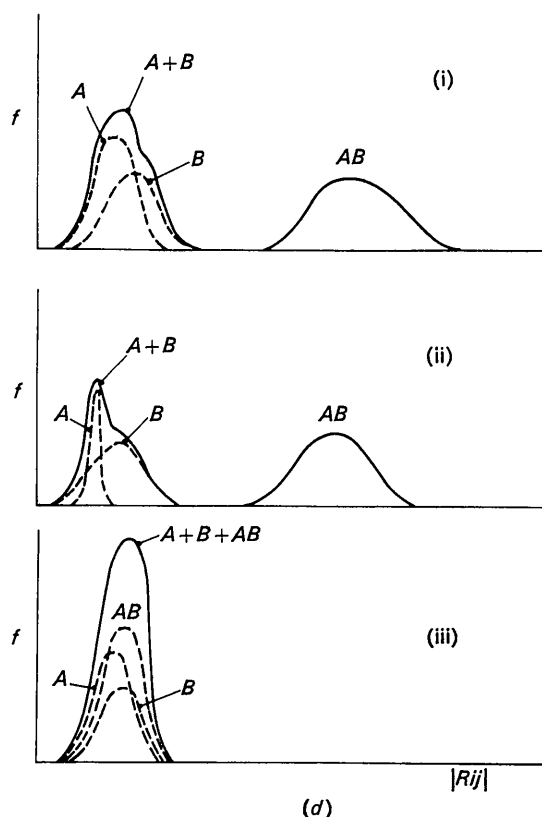


Fig. 1. $|R_{ij}|$ frequency distribution for idealized projects. (a) In which all participants derive identical values. $|R_{ij}|$ is a delta function at $R=0$. (b) In which all participants have marginally small errors of different origin but distributed over the group. (c) As in (b) but one participant involves appreciably greater error than the others. (d) The total group of participants involves two sub-groups, A and B . The various possibilities are illustrated in (i), (ii) and (iii), see text. (i) and (ii) Sub-groups A and B are individually concordant with different spreads but A and B are discordant. (iii) Not only are A and B internally concordant but A and B are now mutually concordant.

(d) Let us elaborate our model one step further. Imagine that in the total group of participants, there are two sub-groups, distinguished by an operational difference *e.g.* (say) one sub-group used Cu and the other Mo radiation. The resultant distribution could be as represented in Fig. 1(d), (i) or (ii) or (iii), depending on the concordancy of the sub-groups *A* and *B*. In (i) or (ii) the sub-groups are internally concordant, *A* being more so than *B* in (ii). The discordancies between *A* and *B* sub-groups are reflected in the *AB* peak. In (iii), the sub-groups are not only internally concordant but they are also mutually concordant. *AB* interactions are therefore not readily distinguished in the peaks due to *A* and *B*.

The idealized situations treated above are sufficient to introduce some of the possibilities for the $|R_{ij}|$ frequency distribution. However it is useful to mention here two other features relevant to our later discussion. The first deals with the question of 'outliers' *i.e.* sets in a project which differ markedly from the consistent majority.

The problem of 'outliers'

Let us imagine that we have one experimentalist, *a*, who can obtain nearly correct values in an absolute sense, *n*-2 average experimentalists who contribute the majority of the data and one other, *o*, who has made a gross slip in setting up his experiment. If we had a set of absolute values, *s*, from a superior independent technique, then the indices $|R_{is}|$ *i.e.* of each set relative to the set *s* would be as in Fig. 2(a). However under normal circumstances, the absolute set, *a*, can neither be obtained nor recognized as such and the resultant frequency distribution, derived in the usual fashion will be as in Fig. 2(b). From the evidence of this distribution above, there is nothing to indicate that the 'outliers' *a* and *o* differ in type. But additional evidence can distinguish a 'good' from a 'bad' outlier. If, for set *o*, it is possible to adduce evidence showing some systematic trend with a particular adjustment or variable and if this trend is revealed and is physically reasonable, there is good reason to treat *o* as a bad 'outlier' and remove it from the derivation of a concordant set. Retention of a sample, established as grossly discordant due to a recognizable cause, is not a useful technique and hinders consideration of the causes of deviations in the main sample. If, as in the case of set *a*, no substantial cause for its removal is shown, it must be retained. However, even here, its being laid aside temporarily may assist in the close investigation of the main group.

Relationships between projects

When the range of variable error-sources is limited, as in a type 2 project, the $|R_{ij}|$ distribution, reflecting this condition, may have the appearance of good concordancy, *e.g.* as in Fig. 3 (type 2). When previously stabilised error-sources become variable error-sources

as in a type 3 project, the $|R_{ij}|$ distribution is correspondingly affected, the peak position moving to a higher *R* value and the peak shape broadening, (Fig. 3, type 3).

The possible relationships between the distributions for projects of different types are of interest. Perhaps this situation can be placed in a clearer perspective by reference to a related group of simplified projects which, by their two-dimensional nature, permit diagrammatic illustration (Fig. 4.) The change in dimensions does not influence the generality of conclusions and the considerations of 'distance' derived apply also to the multi-dimensional cases which concern us in practice.

In this group of projects, *X*, there are two main items to measure H_1 and H_2 . Each experiment is represented by a point corresponding to the two values derived. *T* represents the true or absolute values. The results for the three projects are indicated by different symbols and, for convenience, are outlined as regions 1, 2 and 3 in Fig. 4. The mean values of the measured data for the three projects are given by μ_1 , μ_2 and μ_3 . *I* is the individual experiment which is also a member of all three projects. The $|R_{ij}|$ distributions arising from the projects X_1 , X_2 and X_3 are similar in shape and location to those in Fig. 3, each providing a measure of the internal concordancy of the individual project.

The numerical relation between R_1 , R_2 and R_3 is at present indeterminate. If the error-sources are indeed independent then the probable combination is

$$(R_1^2 + R_2^2)^{\frac{1}{2}} = R_3 \quad (A)$$

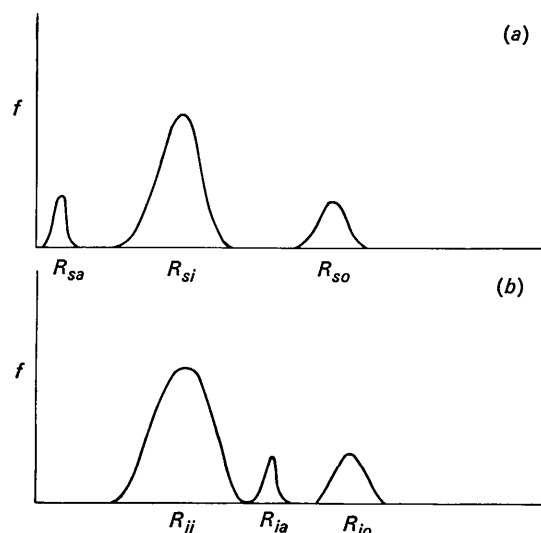


Fig. 2. (a) The frequency distribution $|R_{ij}|$ obtained relative to a set of absolute values *s*, if known as such, where we are dealing with a group consisting of *n*-2 participants of average performance, one *a* who can obtain near correct values in an absolute sense and another *o* who has made a slip in setting up his experiment. (b) The frequency distribution derived in practice for this group when we have no independent means of establishing the absolute values.

If however the error-sources are not independent, the relationship may tend more to a linear relation.

$$R_1 + R_2 = R_3 \quad (B)$$

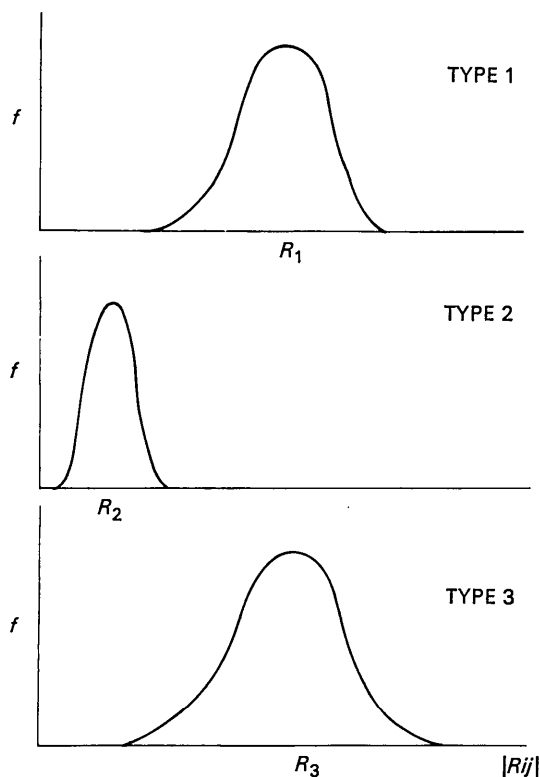


Fig. 3. $|R_{ij}|$ distributions for projects of type 1, 2 and 3 as specified in Table 1.

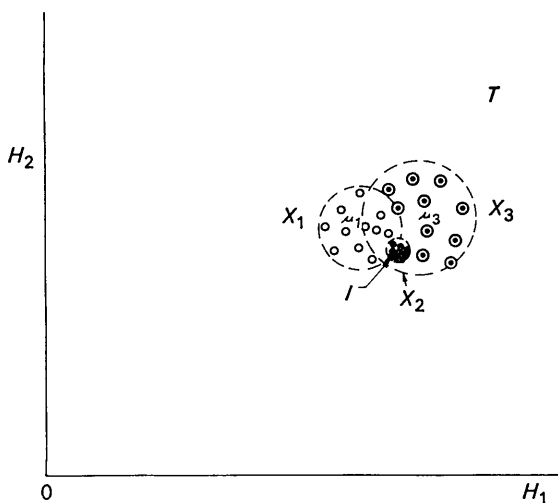


Fig. 4. The relationship between projects of type 1, 2 and 3 illustrated by reference to simplified projects X_1 , X_2 and X_3 involving the measurement of only two items H_1 and H_2 . The mean values respectively of the three projects are μ_1 , μ_2 and μ_3 . T represents the true or absolute values. I is an individual experiment which, for discussion, is visualized as being a member of all three projects. The $|R_{ij}|$ distributions from X_1 , X_2 and X_3 are of the form of those in Fig. 3.

With no time to elaborate the significances of the various possibilities inherent in the viewpoint shown in Fig. 4, we may, however, usefully indicate a few points.

(i) Corrections can be viewed as moving an individual experiment or group of experiments nearer to T , the set of correct values on an absolute basis.

(ii) If there existed only the one experiment I , Fig. 4 stresses the difficulty of giving a measure of the accuracy of such individual experiments from internal evidence alone.

(iii) We may arrive at misleading ideas concerning accuracy if we merely look at the concordancy of F -values derived from projects where certain error-sources are stabilized.

(vi) It seems more probable that mean F -values derived from a type 3 project will approach more closely to the true values even though their precision, as judged by internal concordancy, is poorer than for those derived from projects of type 1 or 2.

(v) $|R_{ij}|$ distributions can provide a guide as to how future projects improve in technique and in terms of numerical results.

(vi) If we are to improve accuracy we must consider all aspects of our experimental techniques and derive corrections so that the $|R_{ij}|$ peak moves towards a zero value not only in one type of project but in all. Hence the importance of establishing all possible error-sources and arranging for their eradication or correction.

(vii) In any practical case, we cannot tell how near to T we are. Although depicted as distant from X_1 (and X_2) in Fig. 4, it may in practice be within X_3 . It is in this respect that an independent set of values from some other experimental technique or soundly-based theoretical calculations may be helpful.

Actual projects

There have been two main projects. The A.C.A. project involved 7 persons who measured ~ 30 orders for CaF_2 , with Mo radiation, each using the same crystal. This was therefore a type 2 project.

The I.U.Cr. project (Abrahams, Hamilton & Mathieson, 1969) involves 16 participants who each measured a different crystal of D(+)-tartaric acid, recording ~ 330 orders. Participants used their own normal procedure and preferred radiation. Data with both Cu and Mo radiations were supplied. This project is therefore of type 3.*

You will all be fully conversant with the contents of the participants report on the A.C.A. project (Abrahams *et al.*, 1967) and there is little point in going over the material presented there. There is however one aspect which may warrant brief comment. As mentioned above, the process of statistical analysis of the

* The anonymity of each participant in both projects was preserved by their being allocated a number. The key to identification was not published and is, for both projects, unknown to the author of the present paper.

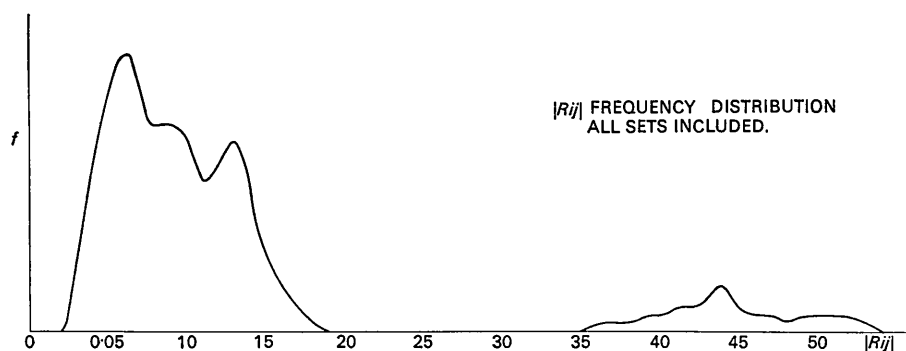
data is of considerable importance in respect of the inferences or conclusions which may be reached. This point can be illustrated by reference to another analysis of the *published data* in the A.C.A. report. This analysis has been carried through by Mackenzie & Maslen (1968). With the author's permission, I will mention

here one feature to illustrate the effect of alteration of viewpoint in analysis. The $|R_{ij}|$ array derived from the original report is shown in Table 2(a). In the analysis of Mackenzie & Maslen, the derivation of a concordant group permitted more detailed diagnosis of individual error-sources. It is not my intention to outline

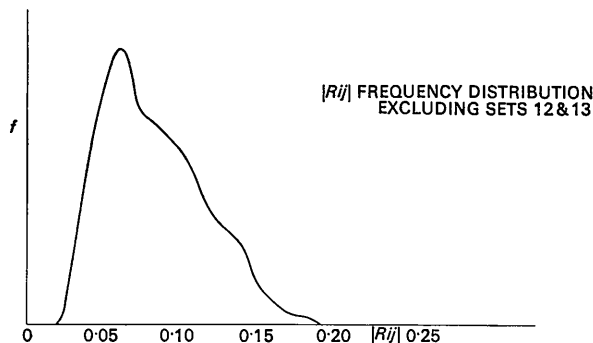
Table 2

(a)							
	1	2	3	4	5	6	7
1	—	0.036	0.055	0.037	0.034	0.062	0.027
2	0.036	—	0.029	0.014	0.011	0.049	0.017
3	0.055	0.029	—	0.032	0.027	0.051	0.029
4	0.037	0.014	0.032	—	0.010	0.053	0.012
5	0.034	0.011	0.027	0.010	—	0.051	0.013
6	0.062	0.049	0.051	0.053	0.051	—	0.047
7	0.027	0.017	0.029	0.012	0.013	0.047	—

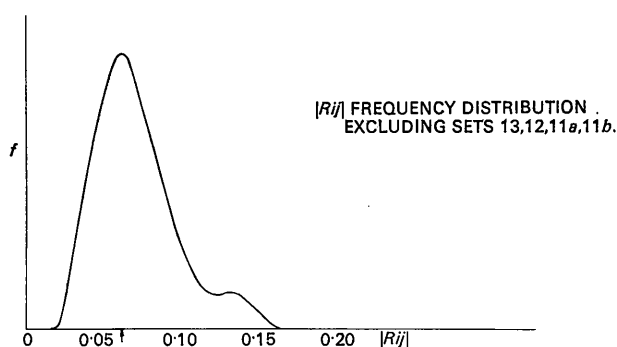
(b)								
	6	3	4	1	2	5	7	
6	—	0.047	0.050	0.052	0.045	0.049	0.049	Fixed crystal, fixed counter, 4-circle. ω -scan, equi-inclination. ω -scan, equi-inclination + crystal monochromator.
3	0.047	—	0.025	0.026	0.024	0.022	0.020	
4	0.050	0.025	—	0.006	0.014	0.010	0.009	
1	0.052	0.026	0.006	—	0.011	0.011	0.009	$\omega/2\theta$ -scan, 4-circle.
2	0.045	0.024	0.014	0.011	—	0.011	0.012	
5	0.049	0.022	0.010	0.011	0.011	—	0.009	
7	0.049	0.022	0.009	0.009	0.012	0.009	—	



(a)



(b)



(c)

Fig. 5. The $|R_{ij}|$ distribution for the I.U.Cr. project data. (a) for the complete group of 17 sets, (b) after removal of sets 12 and 13 (c) after removal of sets 11a, 11b, 12 and 13. The $|R_{ij}|$ array from which the distributions are derived is given in Table 3

these here but it is perhaps useful to show the $|R_{ij}|$ array obtained on deletion of only the 3 lowest angle terms, Table 2(b). Here, the instrumental types have been ordered in a definite sequence so that we have a ranking of techniques. The implications of these results will be obvious.

In the case of the I.U.Cr. project, a full report is under way and it is not my function here to present the mass of material available. However, some interpretations of the basic data, to illustrate the use of the $|R_{ij}|$ distribution are suitable on the present occasion.

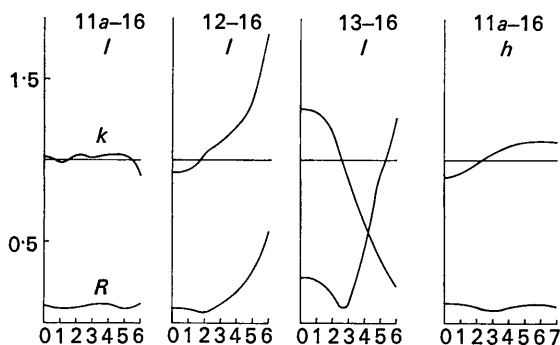


Fig. 6. Test of sets 11a, 12 and 13 relative to a 'typical' set, 16, against index l .

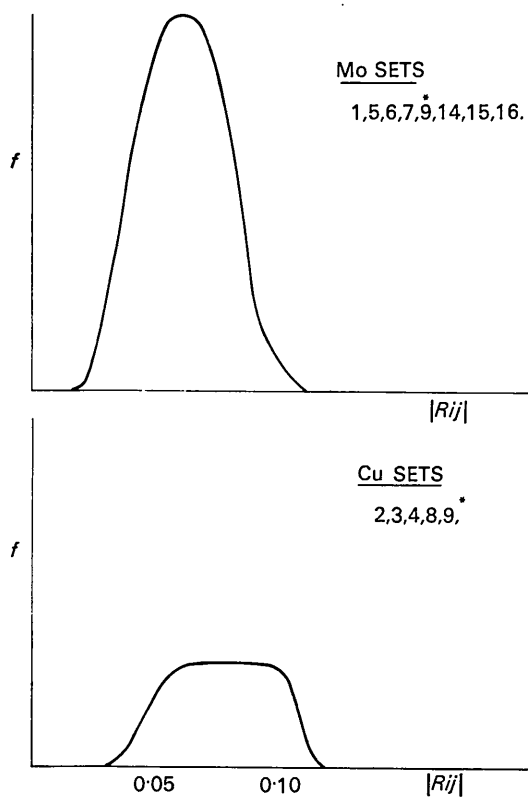


Fig. 7. The $|R_{ij}|$ distributions for the I.U.Cr. subgroups corresponding to the sets in Fig. 5(c) involving one radiation (i) Mo (ii) Cu.

The frequency distribution for the complete set (using a moving average) is given in Fig. 5(a). Set 13 is an example of a gross outlier while 12 also shows a similar trend. Removal of sets 12 and 13 led to the result in Fig. 5(b) while further extraction of sets 11a and 11b yielded Fig. 5(c). The corresponding peak $|R_{ij}|$ value is then $\sim 6\%$. That sets 13 and 12 are indeed subject to gross error is demonstrated by supplementary evidence in which the data for sets 12 and 13 are plotted relative to an arbitrarily chosen representative set 16 against index l as variable (Fig. 6.) The smooth systematic trend with layer index for 13 and 12 and the fact that sets 11, 12 and 13 were recorded on equi-inclination instruments virtually establishes that the gross error-source is probably associated with the equi-inclination angle.

It is of interest to record that the sub-groups derived from Mo and Cu radiations respectively reveal distributions which differ in peak position and shape (Fig. 7). The R peak for Mo is both at a lower value and more compact than that for Cu – as one might expect. However, it is evident that the two sub-groups are of the type depicted in Fig. 1(a) (iii) and are not grossly discordant.

From the analysis-of-variance for trends with intensity, d^* , and indices h , k and l , it is found that there is a sub-group of sets which does not show significant trends with any of these variables. For this sub-group, the $|R_{ij}|$ distribution (data not rescaled) has a peak value of $\sim 4\%$.

Some implications

It should now be evident that comparison of type 2 and type 3 projects indicates the influence of the crystal as an error-source and can provide some guide as to its magnitude.

Obviously we are proposing here to compare results for different materials, tartaric acid and CaF_2 . However, it is of interest to note that the second component of the A.C.A. project provides some sort of tie-point between the two projects. This evidence is limited to only one row of the $|R_{ij}|$ array. Within this limitation, the mean value ($\equiv |R_{ij}|$ peak) lies essentially in the same range, 6%, as the peak value of the main set of the I.U.Cr. project, Fig. 7(i) so that the projects can not be regarded as totally dissimilar in regard to the totality of error-source contributions.

1. This being the case we may allow ourselves an extrapolation in comparing the two projects in respect of the full sets and also the restricted concordant sets.

(i) The difference in peak $|R_{ij}|$ values for the full sets are as follows depending on whether combination (A) and (B) applies:

$$\text{If (A), then } (0.062 - 0.04)^{\pm} = 0.045 .$$

$$\text{If (B), then } 0.06 - 0.04 = 0.02 .$$

(ii) For the restricted concordant sets

$$\text{If (A), then } (0.042 - 0.01)^{\pm} = 0.04 .$$

$$\text{If (B), then } 0.04 - 0.01 = 0.03$$

There is no intention here of attaching too much weight to these numbers, only of suggesting an order of magnitude. They do indicate however that, apart from other error-sources the crystal contributes $\sim 3 \pm 1\%$ on this basis of assessment. One must remember that $|R_{ij}|$ gives an overall assessment of the fit. Obviously individual reflexions may be subject to wider errors.

It is of interest to note that an overall figure of 4% was suggested for NaCl crystals by Abrahams (Arndt & Willis, 1966). It was also suggested there that this total error be regarded as shared equally between diffractometer and crystal. In this case, the error contribution from the crystal would be 3% [if based on (A)] or 2% if based on (B).

2. Even taking into account the different bases of the two projects, it is notable that, in each, the sets form a concordant sub-group and the others markedly less concordant. We can therefore appreciate that these latter may be due to inclusion of an experimental technique which is not suited to the purpose of attaining accurate intensities or to an instrument type whose adjustment must be recognised as crucial and act accordingly.

3. Out of both the re-assessment of the A.C.A. project (Mackenzie & Maslen, 1968) and the I.U.Cr. project, certain inferences may be made concerning operational details in the measurement of F -values but these I will leave to appear in their proper place. The possibilities of serious error in the use of equi-inclination devices are evident in Tables 2 and 3. The values in Table 2(b) draw particular attention to the importance of crystal monochromatization for equi-inclination devices. Although this conclusion may be in line with earlier comments, we have here some striking indications to establish the point.

4. I come now to a matter which may be of more practical interest and concern – namely the significance of $|R_{ij}|$ considered as an analysis of pairs of experiments.

The results from the I.U.Cr. project are most relevant since, being of type 3, they approximate closely to practical situations. If we disregard the grosser features attributable to mis-settings of μ and consider the results for the main group, Fig. 5(c), you will note that, if the I.U.Cr. participants are not untypical, Fig. 5(c) suggests that if two crystallographers carry measurements of F -values by their usual procedure, it is likely that the difference, judged by R_{ij} will be on the average $\sim 6\%$. It is unlikely that they will fit better than 3% or worse than 10%. This applies to crystallographers measuring different crystals and using their own normal technique – a situation applicable to the majority of structure studies. The results suggest an explanation of a matter which has been commented on occasionally in the past. Namely that it is easier to obtain a good match between a set of *calculated* and a set of *experimental structure factors* as assessed by $|R_{oc}|$ than between *different sets* of *experimental structure factors*,

Table 3. $|R_{ij}|$ and $|R_{ij\mu}|$ values – I.U.Cr. project

	13	12	11b	11a	14	15	16	2	3	4	8	1	5	6	7	9	10	λ	$R_{ij\mu}$
13	—	0.497	0.359	0.416	0.430	0.383	0.443	0.482	0.434	0.480	0.460	0.401	0.435	0.520	0.448	0.406	0.508	Mo	0.478
12	0.497	—	0.164	0.126	0.138	0.151	0.123	0.135	0.130	0.136	0.130	0.128	0.121	0.140	0.120	0.139	0.162	Mo	0.141
11b	0.359	0.164	—	0.143	0.140	0.093	0.133	0.153	0.102	0.171	0.156	0.128	0.135	0.114	0.128	0.100	0.116	Mo	0.092
11a	0.416	0.126	0.143	—	0.086	0.118	0.104	0.098	0.099	0.093	0.113	0.111	0.100	0.116	0.096	0.109	0.125	Mo	0.093
14	0.430	0.138	0.140	0.086	—	0.098	0.071	0.106	0.059	0.099	0.092	0.085	0.070	0.069	0.065	0.076	0.088	Mo	0.062
15	0.383	0.151	0.093	0.118	0.098	—	0.075	0.126	0.066	0.135	0.110	0.071	0.069	0.042	0.085	0.066	0.034	Mo	0.058
16	0.443	0.123	0.133	0.104	0.071	0.075	—	0.070	0.040	0.087	0.051	0.044	0.038	0.048	0.034	0.060	0.071	Mo	0.048
2	0.482	0.135	0.153	0.098	0.106	0.126	0.070	—	0.090	0.047	0.061	0.046	0.084	0.120	0.060	0.091	0.145	Cu	0.084
3	0.434	0.130	0.102	0.099	0.059	0.066	0.040	0.090	—	0.097	0.067	0.062	0.037	0.042	0.043	0.056	0.048	Cu	0.031
4	0.480	0.136	0.171	0.093	0.099	0.135	0.087	0.047	0.097	—	0.068	0.072	0.086	0.128	0.071	0.099	0.137	Cu	0.091
8	0.460	0.130	0.156	0.113	0.092	0.110	0.051	0.061	0.067	0.068	—	0.063	0.060	0.074	0.046	0.073	0.094	Cu	0.071
1	0.401	0.128	0.128	0.111	0.085	0.071	0.044	0.046	0.062	0.072	0.063	—	0.048	0.072	0.041	0.062	0.083	Mo	0.053
5	0.435	0.121	0.135	0.100	0.070	0.069	0.038	0.084	0.037	0.086	0.060	0.048	—	0.051	0.044	0.061	0.057	Mo	0.046
6	0.520	0.140	0.114	0.116	0.069	0.042	0.048	0.120	0.042	0.128	0.074	0.072	0.051	—	0.059	0.052	0.033	Mo	0.040
7	0.448	0.120	0.128	0.096	0.065	0.085	0.034	0.060	0.043	0.071	0.046	0.041	0.044	0.059	—	0.057	0.080	Mo	0.049
9	0.406	0.139	0.100	0.109	0.076	0.066	0.060	0.091	0.056	0.099	0.073	0.062	0.061	0.052	0.057	—	0.051	CuMo	0.049
10	0.508	0.162	0.116	0.125	0.088	0.034	0.071	0.145	0.048	0.137	0.094	0.083	0.057	0.033	0.080	0.051	—	Cu	0.039

equi-inclination

4-circle

r-beam

(?)

as assessed by $|R_{ij}|$. If the spread of experimental values is as large as is indicated in Fig. 5(c) then it is clear that the flexibility of the multi-variable theoretical structure factor calculations has tended to absorb and obscure many of the errors which exist in the experimental F -values.

This conclusion, and the results shown in Fig. 5(c), are rather disturbing in relation to structure analysis, particularly the study of fine details. It raises doubts as to the limits to which a refinement against a theoretical atomic model should be pushed in the case of measured F -values derived in a single isolated experiment. It further raises the question of the validity of many structural conclusions concerning atomic and vibration parameters derived from such studies. In particular, it stresses the need for measured F -values to be based on as wide a range of experimental techniques as possible.*

I wish to express my warmest appreciation for the valuable advice and guidance on statistical matters given by my colleagues Drs. J. K. Mackenzie and W. Maslen and also for permission to quote from their unpublished work. I am also deeply grateful to my colleague Dr D. A. Wright for careful critical reading of the manuscript and for his contributions to its clarification. I should also like to express appreciation to my Commission colleagues, Drs S. C. Abrahams, V. W. Maslen and W. C. Hamilton, for numerical values derived in the analysis of the I.U.Cr. project and acknowledge their agreement to the use of such data prior to full publication of the project report.

References

- ABRAHAMS, S. C. (1964). *Acta Cryst.* **17**, 1190.
 ABRAHAMS, S. C., ALEXANDER, L. E., FURNAS, T. C., HAMILTON, W. C., LADELL, J., OKAYA, Y., YOUNG, R. A. & ZALKIN, A. (1967). *Acta Cryst.* **22**, 1.
 ABRAHAMS, S. C. & BERNSTEIN, J. L. (1965). *Acta Cryst.* **18**, 926.
 ABRAHAMS, S. C. & HAMILTON, W. C. (1969). *Acta Cryst.* I. U. Cr. Single Crystal Intensity Project. Part II. To be published.
 ABRAHAMS, S. C., HAMILTON, W. C. & MATHIESON, A. McL. (1969). *Acta Cryst.* I. U. Cr. Single Crystal Intensity Project. Part I. To be published.
 ARNDT, U. W. & WILLIS, B. T. M. (1966). *Single Crystal Diffractometry*, p. 298. Cambridge Univ. Press.
 FISHER, R. A. (1947). *The Design of Experiments*. London: Oliver & Boyd.
 MACKENZIE, J. K. & MASLEN, V. W. (1968). *Acta Cryst.* **A 25**, 629.
 MASLEN, V. W. (1967). *Acta Cryst.* **23**, 907.
 SCHOMAKER, V. (1965). *Trans. Amer. Cryst. Ass.* **1**, p. 30.

* I have dealt with matters relating to *measurement* and have avoided specific reference to comparisons with theoretical F -values. The data in each set of the I.U.Cr. project has been put through the 'black box' of a standard Busing-Levy least-squares refinement. The results of this manipulation will be reported subsequently (Abrahams & Hamilton, 1969).

TOGAWA, S. (1964). *J. Phys. Soc. Japan*, **19**, 1696.

WEISS, A., WITTE, H. & WÖLFEL, E. (1957). *Z. phys. Chem.* **10**, 98.

DISCUSSION

YOUNG: The importance of the inferences you derive depends on people in the I.U.Cr. project using their normal procedure. I wonder if, in fact, normal procedures were used for the absorption correction. The crystals we used were of natural shape and somewhat larger than our normal selection. However, I passed on the data assuming that a program to deal with this would be available, and that suitable absorption corrections would be applied. In this respect, our contribution was therefore not normal and my experience may not be unique in this respect. Thus, in the draft report on the I.U.Cr. project, there are indications in the d^* trends of a possible correlation with absorption. Therefore the possibility is that the situation is not quite as bad as it at first appears in the report. Probably we do better than this normally.

MATHIESON: I will allow that these represent extreme conditions. However, they do represent quite closely the type of results that structure analysts turn in. As I mentioned, I tried to extract out of the two projects some indication of the crystal. One can see that dropping the first three reflexions of the A.C.A. project leads to a rather different picture with a figure of ~ 1 to $1\frac{1}{2}\%$. This represents the reproducibility over a number of diffractometers or in other words represents the error effect of the diffractometer. What one derives from the I.U.Cr. project is a whole combination of errors available, or likely to occur, in normal situations, I have tried to extract a meaningful number for the crystal as error source and the value of 3–4%, arrived at, does in fact accord in some sense with the value concerning NaCl referred to in *Single Crystal Diffractometry* (Arndt & Willis, 1966).

ABRAHAMS: I would like to make a comment on the difficulty of making a completely sensible analysis of this sort of project. The point I would like to stress relates to the results presented in Part II of the I.U.Cr. Project Report (To be published in *Acta Cryst.*). Thus if one asked which of the sets disagreed least with each other, we end up with five sets. If we then put these through a standard least-squares procedure, we derive the positional and thermal parameters for each set, with the standard deviation derived for each set. One can see the results in the draft of Part II. The spread of values relative to the computed is surprisingly large. This illustrates that however closely a group of experiments may fit on one basis, they may well show much wider deviations when considered from a different viewpoint.

BRAIBANTI: Were the temperatures of the individual laboratories taken into account?

MATHIESON: No. There were essentially the situations that applied in the laboratory of the participants. We should realize that this is indeed a true comparison of results of the type that would be reported in *Acta Cryst.* The principal point that I would like to bring out is that an individual experiment, *carried out once*, is a unique experience and is therefore difficult to assess on its own.

FURNAS: How many of the sixteen participants who turned in data for the I.U.Cr. project felt, as we did, that the crystals were not suitable and not worth spending effort on? In fact were the actual contributors representative only of those willing to go along with the game, *i.e.* be cooperative?

MATHIESON: This is a serious and important factor. Earlier we had looked at many materials, tested them and discarded them. This particular choice, tartaric acid, seemed reasonable at the time, and only subsequently did we realize that there were subtle features in terms of the stability of these crystals. (Professor A. F. Peerdeman has made a close study of the crystals and appears to have an explanation for their particular behaviour.) In fact, the situation has parallels in the A.C.A. project where an early choice of NaCl was changed to CaF₂. Even with this choice there are problems associated with extinction. For phase II of the I.U.Cr. Project, aimed to commence in 1969, we have continued our study of possible materials; all have defects, as for instance X-radiation damage. We have not arrived at a sensible practicable choice of crystal material to pass around. This all sounds rather negative but illustrates the problems besetting choice of a standard. One should realize that a normal structure analysis involves selection of one or two crystals, data are collected relatively rapidly and data manipulation proceeds. One never really sits around watching the crystal for days or weeks. The situation in a project, by contrast, is a very severe test of crystal stability. We would be very grateful for reasonable suggestions, *from first-hand experience*, of stable, low-absorption, preferably low-symmetry materials which we could test further for phase II.

We should also recognize as I mentioned earlier, that we are learning slowly about projects and their design. I rather think that from our experience we can lay down a basis design for phase II rather better than when the project was initiated originally.

SCHOMAKER: May I offer a comment? If you found this material, you could reduce the R_{ij} peak for type 3 project to that of the type 2 project.

MILLEDGE: Amplifying the earlier comment by R. A. Young, I wonder if the initial part of the project might have had a different basis? That is, by inviting participants to provide their raw data and then have the appropriate corrections applied by the project analysts. It is just possible that correction procedures by the individual participants might in some cases have been less than satisfactory and, in fact, the raw data were better than they appeared on analysis of the individually derived F^2 values. Is it still possible to obtain the raw data and treat them in the manner proposed?

ABRAHAMS: In principle, I completely agree but in practice, this would involve setting up comprehensive correction programs for the various possible types of diffractometer geometry. I am not sure that it would be possible for one individual to do adequate justice to such a procedure. It is just possible that this might introduce errors of other sorts.

MATHIESON: Such a situation would be slightly synthetic. One basic thought associated with the I.U.Cr. project was this: How do experimental results compare, done in different laboratories, in different parts of the world, each

experimentalist proceeding with his usual procedure and applying his corrections again in his usual manner. This is an important aspect of the results. Apart from the outliers, which I think we can detect, even those sets which are acceptable are still in the high range of $\sim 6\%$. These represent the real situation, the structure analyses which are published in the literature.

MILLEDGE: Yes, I agree, but it might still make a valuable contribution to the subject to allow interested laboratories to work over the raw data.

ZACHARIASEN: I would like to make some comments on the A.C.A. project – in which I did not participate, being a lone operator. When the F^2 values were published in *Acta Cryst.*, I looked at them, being interested in extinction at the time and some slight test showed that extinction in the primary standard CaF₂ crystal was really enormous. I had been looking around for crystals with high extinction and this was it. For example, the 220 reflexion had an F_0^2 approximately 14% of F_c^2 . Obviously, whatever success you might say that project had was entirely dependent on the use of one crystal. If there had been seven different crystals, the results, in my opinion, would have had no sense whatever.

Another thing I would like to emphasize since I have run into it a number of times. First, however, let me say that I do not use an automatic diffractometer but collect my data by hand one at a time, so that if anything appears which is out of the ordinary, I see it straight away. To come to the point: if you spend a lot of time collecting data with one radiation *e.g.* Mo, I would urge that you also collect at least 10% of your reflexions using a different radiation *e.g.* Cu. For example, when visiting Abrahams's laboratory, I asked if I might borrow the CaF₂ primary crystal but it had been lost. However there was a secondary standard which was within 1% of the radius of the primary standard. So I collected a complete set of measurements on all reflexions up to 90°. In this specimen, there was also high extinction but slightly smaller than in the primary

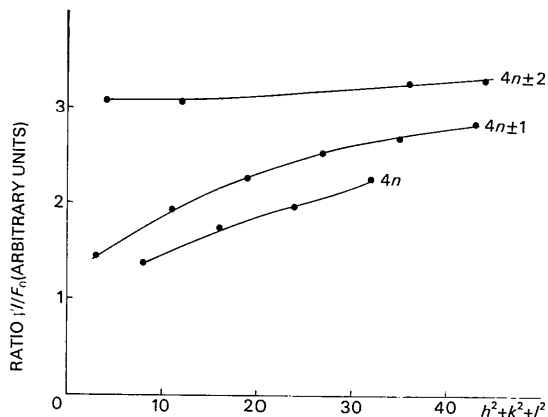


Fig. 8. Extinction effects for the primary crystal of CaF₂ used in the ACA project. The ratio I/I_0 is plotted against $h^2 + k^2 + l^2$. Values of the mean intensities I , which are not corrected for extinction, were obtained from the ACA report. Absolute structure factors F_0 are from Togawa (1964) for the strong and medium reflexions ($h+k+l=4n$ and $4n \pm 1$, respectively) and from Weiss, Witte & Wölfel (1957) for the weak reflexions ($4n \pm 2$).

standard. I then transferred to Cu radiation. From my approximate formula I knew the extinction should be less since $\mu R \approx 6.4$ for Cu and 0.6 for Mo. Sure enough the extinction was less but in fact much smaller than I predicted. Obviously there was some other factor which I believe was the Borrmann effect which, although smaller for a sphere than for a plate, nevertheless cannot be ignored in crystals exhibiting high extinction.

MATHIESON: May I add to that. Fig. 8 (for which I am indebted to Dr V. Maslen) shows the CaF_2 data for the A.C.A. project compared with Togawa's experimental results. The data are sub-divided into groups (a) with $h^2 + k^2 + l^2 = 4n \pm 1$ where the structure factor is a function of Ca-F , (b) $h^2 + k^2 + l^2 = 4n + 2$ where $\text{S.F.} \equiv \text{Ca}$, and (c) $h^2 + k^2 + l^2 = 4n$ where $\text{S.F.} \equiv \text{Ca} + 2\text{F}$. The influence of extinction on the different sub-groups, and hence on the different intensity levels, is fairly obvious.

FURNAS: There are many different computer programs available. Could one standard set of data be distributed in order to gain some idea as to what differences appear in final atomic dimensions as a result of variation in program type?

HAMILTON: The Commission on Crystallographic Computing of the Union is at present assembling a set of test calculations for least squares, Fourier, bond distances and angles and probably absorption. With this, artificial data will be supplied and it is proposed to provide these for several space groups.

HAHN: While the average R_{ij} is $\sim 6\%$, you mentioned that five more concordant sets had been refined by least squares. What was the lowest R -factor for observed *versus* calculated F in this group?

ABRAHAMS: The various values are given in Table 4 of Part II of the Report on the I.U.Cr. Project. The lowest value is 0.034.

HERBSTSTEIN: Were these sets put on an absolute or quasi-absolute scale? It seems to me that we have reached a situation where it would be possible and advisable to do so. Dr Chipman indicated in session G.2 that it could be done for powders. If it could be done for single crystals in a reliable manner, it would remove one source of error.

MATHIESON: None of the sets were placed on an absolute scale experimentally. They were simply placed on a common scale after being gathered together.

JEFFERY: (a) It is a pity that, due to a fire, we were not, in fact, able to supply film data on tartaric acid. It would appear from the figures given that our partial results fitted in reasonably.

(b) One point which was made evident in adjusting our equi-inclination camera was the original error in the cell parameter, a , given for tartaric acid. In this respect, the film technique illustrates its self-checking potentiality.

ABRAHAMS: The original cell parameters circulated were very precise and highly inaccurate. The fact is that the specimen crystal consisted of mis-orientated blocks. As a result, measurements on the Bond apparatus of high-order reflexions on the $+2\theta$ region were in fact carried out on different blocks of the crystal. These results reflected one

of the problems, alluded to by Dr Furnas, that the crystals had gross mosaic spread, but not in all cases, there was a variation. This sort of thing has, of course, worried us for quite a long time. In our search for a standard crystal, we have looked at many with a jaundiced eye and in all cases with justification.

SCHOMAKER TO R. J. WEISS: Is the big difference between the momentum distribution for graphite and diamond relative to that calculated for the single atom, in accord with the small effects described earlier by Dawson which require to fit the intensities. Are they consistent in respect of order of magnitude or not?

WEISS: Yes.

EWALD: I would like to ask – is this a concluded project of the I.U.Cr. or is it going on? My impression is that it is a preliminary investigation of how to investigate the question. So my view is that it should be continued under all circumstances. It shows the value of cooperation between various laboratories all over the world.

MATHIESON: This is exactly our viewpoint. We are regarding this as essentially phase I of the project. We are learning how to tackle the job. What we think we will do is to have a rest and try to find a suitable material, one which is acceptable and will stand up to ambient conditions in various laboratories. We intend to repeat the process later with perhaps a different design of analysis. In other words, incorporate what we have learned. I think, in fact, we have learned a great deal. Thus, under normal circumstances, the stationary-crystal stationary-counter procedure is out if we are aiming for high accuracy, the equi-inclination method needs careful operation, Mo radiation is better than Cu for certain operations and there are other features which will come out in the detailed reports.

You may say we knew all these things before but in actual fact we did not have any numbers attached. Now we have. However we do intend to proceed to phase II.

HIRSHFELD: May I suggest that if the project is to continue, it might gain from a shift of emphasis. Rather than have participants measuring a very large number of reflexions, select a smaller number and invite the participants not only to do their very best but also to supply some measure of their personal accuracy on each measurement *e.g.* a standard deviation. This would permit subsequent comparisons on the basis of weighted R indices.

MATHIESON: I think this is an excellent suggestion and we should keep it in mind. On the other hand, we should realize that we are dealing with people and not automatic machines. We did have a large number offering originally to take part in the project. The eventual number proved to be sixteen. We must take into account that all these efforts are voluntary and that we cannot twist arms too much.

It is perhaps an appropriate point at which to say how much we appreciate the contributions of all those participants who sent in results. I do not think we can repeat our thanks too often. It is only with their assistance that the analysis of the project is feasible and it is only with their data that we can show if we are doing better next time or whether we are not improving at all. I would like to say how deeply grateful we are to all who contributed to phase I of the I.U.Cr. Project.