Acta Cryst. (1966). **21**, 648

New Statistical Tests for Distinguishing between Centrosymmetric and Non-Centrosymmetric Structures*

BY R. SRINIVASAN AND T. SRIKRISHNAN

Centre of Advanced Study in Physics, University of Madras, Madras 25, India

(*Received* 30 *August* 1965 *and in revised form* 5 *December* 1965)

Some new statistical tests for distinguishing between centrosymmetric and non-centrosymmetric structures are described. The tests are based on the possibility of generation of a random set of intensities from the data available for a given crystal such as to afford a pair of independent variables. The probability distributions and other statistical parameters connected with such a pair are given and are also illustrated on practical cases.

Introduction

We describe here some new statistical tests for distinguishing between centrosymmetric (centric) and non-centrosymmetric (acentric) structures. They are based on the distribution and other statistical parameters connected with two independent variables each one having intensity distribution characteristic of a centric or an acentric structure as the case may be. The applicability of the results in actual practice to data of a single crystal is based on the idea of generating a second sample by a random permutation of the available sample values. Although it might appear that such a procedure is rather artificial, the tests nevertheless have their own distinctive features of advantage which indicate that they would prove useful as statistical criteria.

Distributions connected with gamma variables

Suppose we have the normalized intensities $z_1 = I_1/\langle I_1 \rangle$ of a given crystal. One could imagine another set of variables z_2 obtained by a random permutation of the given values of z_1. If the permutation is really random, it can be assumed that the two variables z_1 and z_2 may be treated as a pair of independent variables. It is possible to work out, quite generally, the statistics of such a pair, and these afford us practical criteria for use as tests for centrosymmetry.

Fortunately, the probability distributions connected with such a pair can be readily derived by the use of certain standard results. Thus, it was pointed out in an earlier communication (Srinivasan & Subramanian, 1964) that the variable z for the acentric case and the variable $z' = z/2$ for a centric case are characterized by gamma variables with parameters 1 and $\frac{1}{2}$ respectively. There also exist a number of results connected with a pair of independent gamma variables. For instance the sum of two such variables can be shown to be again a gamma variable and also their quotient has a beta distribution and so on. These results are available in standard works (*e.g.* Weatherburn, 1961, pp. 146–63) from which the results pertaining to our case can be readily deduced*. We therefore omit the proofs, and present only the results.

Given a pair of independent gamma variables z_1 and z_2 we define the following variables

$$t = z_1/z_2,\ r = z_1/(z_1+z_2)\ . \tag{1}$$

Note that the quotient t has a range 0 to ∞ while r has the finite range 0 to 1. The distribution of t and r can be shown to be

$$_AP(t)dt = \frac{dt}{(1+t^2)^2} \tag{2a}$$

$$_CP(t)dt = \frac{dt}{\pi t^{\frac{1}{2}}(1+t)} \tag{2b}$$

$$_AP(r)dr = dr \tag{3a}$$

$$_CP(r)dr = \frac{dr}{\pi\sqrt{r(1-r)}} \tag{3b}$$

where we have used subscripts A and C to denote acentric and centric cases respectively. We shall be interested in two more variables, v and u, defined by

$$v = t^{\frac{1}{4}} \text{ and } u = r^{\frac{1}{4}}\ . \tag{4}$$

These involve the structure amplitudes or similar quantities. The distributions for v and u are readily worked out from (2*a*), (2*b*), (3*a*) and (3*b*). They are given by

$$_AP(v)dv = \frac{2v}{(1+v^2)^2}\,dv \tag{5a}$$

$$_CP(v)dv = \frac{2}{\pi} \times \frac{1}{(1+v^2)}\,dv \tag{5b}$$

$$_AP(u)du = 2u\,du \tag{6a}$$

$$_CP(u)du = \frac{2}{\pi\sqrt{1-u^2}}\,du \tag{6b}$$

* Contribution No. 197 from the Centre of Advanced Study in Physics, University of Madras, Madras 25.

* Some of these results have already been obtained earlier in another context (Srinivasan, Subramanian & Ramachandran, 1965). The present paper may, however, be read independently.

Correlation coefficient and other statistical parameters

The applicability of the above results to the pair z_1 and z_2 depends on the assumption that they are independent. It is preferable, in practice, to adopt some simple criterion to check the independence of such a pair. For instance, one can use the usual coefficient of linear correlation, which is defined for any given pair of variables as

$$k=\frac{\langle(x-\langle x\rangle)\rangle\langle(y-\langle y\rangle)\rangle}{[\langle(x-\langle x\rangle)\rangle^2\langle(y-\langle y\rangle)\rangle^2]^{\frac{1}{2}}}. \qquad (7)$$

If k has a value zero or close to zero the two may be taken to be independent. One may take in (7) x and y to be the normalized intensities z_1 and z_2 or the normalized structure amplitudes y_1 and y_2 where $y_1=\sqrt{z_1}$ and $y_2=\sqrt{z_2}$.

It is of interest to point out in this connexion that one may also evaluate a closely related quantity* C_A, defined by

$$C_A=\langle y_1y_2\rangle=\frac{\langle|F_1||F_2|\rangle}{[\langle|F_1|^2\rangle\langle|F_2|^2\rangle]^{\frac{1}{2}}} \qquad (8)$$

$$=\frac{\Sigma|F_1||F_2|}{[\Sigma|F_1|^2\,\Sigma|F_2|^2]^{\frac{1}{2}}}. \qquad (9)$$

It is obvious that if $|F_1|$ and $|F_2|$ are independent, since $\langle|F_1|^2\rangle=\langle|F_2|^2\rangle=\langle|F|^2\rangle$ and $\langle|F_1|\rangle=\langle|F_2|\rangle=\langle|F|\rangle$, (9) is nothing but $\langle|F|\rangle^2/\langle|F|^2\rangle=\varrho$ where ϱ is the well-known ratio of Wilson (1949) which has a value $\pi/4$ for non-centrosymmetric and $2/\pi$ for centrosymmetric cases. This would be an additional test with such a pair and may be checked against the direct calculation of ϱ.

Other statistical parameters

We give two other simple parameters which are based on the mean values of the quotient variables. First we define new variables v' and t' by

$$v'=v \quad \text{for } v\le 1 \qquad (10)$$
$$=1/v \text{ for } v\ge 1$$

and similarly

$$t'=t \quad \text{for } t\le 1 \qquad (11)$$
$$=1/t \text{ for } t\ge 1\,.$$

We now evaluate the expectation values $\langle v'\rangle$ and $\langle t'\rangle$. These are readily worked out from the available distributions $P(v)$ and $P(t)$ and the proofs are omitted. The value $\langle v'\rangle$ reduces to $\frac{1}{2}\pi-1$ for non-centrosymmetric and $(2/\pi)\log_e 2$ for the centrosymmetric case. Similarly $\langle t'\rangle$ reduces to $(2\log_e 2-1)$ and $(4/\pi-1)$ respectively for the acentric and centric cases. The reason why v' and t' instead of v and t are used is that the mean values of these do not exist in certain cases.

While $\langle v'\rangle$ and $\langle t'\rangle$ might themselves be useful, we define two other quantities

$$R'=1-\langle v'\rangle \qquad (12)$$
$$R''=1-\langle t'\rangle \qquad (13)$$

where the use of the symbol R will become clear shortly. Thus, we have

$$R'=1-\langle v'\rangle=\left\langle\frac{||F_1|-|F_2||}{|F_1| \text{ or } |F_2|}\right\rangle$$
$$=\frac{1}{N}\Sigma\left(\frac{||F_1|-|F_2||}{|F_1| \text{ or } |F_2|}\right), \qquad (14)$$

where N is the number of pairs of values used in the calculation and the 'or' in the denominator of (14) means that $|F_1|$ or $|F_2|$, which ever is larger, is to be used for averaging. So, also, it is readily seen that

$$R''=1-\langle t'\rangle=\frac{1}{N}\Sigma\left(\frac{|I_1-I_2|}{I_1 \text{ or } I_2}\right). \qquad (15)$$

The quantities on the right-hand side of (14) and (15) may be interpreted as the mean values of the percentage discrepancies of the pairs, and they resemble closely the conventional reliability index although they are distinctly different from it.

Application of the results as tests for centrosymmetry

The distributions (2), (3), (5) and (6) are shown in Figs.1 to 4. (The points wherever marked pertain to practical tests and will be explained later). It may be noticed that in each case the curves for the centric and acentric cases are quite dissimilar in shape. In particular the functions $P(r)$ and $P(u)$ are interesting. ${}_AP(r)$ and ${}_AP(u)$ are actually straight lines. Note that ${}_AP(r)$ has a constant value of unity. Physically, this only reflects the rather high uniformity in the intensity distribution for a non-centrosymmetric crystal. The two functions $P(r)$ and $P(u)$ have the additional feature that they have a finite range.

Thus, all these functions may be expected to be useful as practical tests for centrosymmetry. They were actually tested in practical cases. These and certain other practical aspects will be described presently.

Tests in practical cases

The functions are all defined in terms of the normalized intensities. Thus, one converts, initially, the available intensities I by dividing them by the local average intensities $\langle I\rangle$ to obtain the set z_1. The second set of values z_2 may be generated by a process of random picking from the first set and pairing them arbitrarily. Tables of random numbers might be helpful in this connexion (*e.g.* Kenney & Keeping, 1954). It may be pointed out here that since the set z_2 is generated from the set z_1, it is possible to deal directly with the actual intensities I_1 and I_2. This, however, demands that $\langle I\rangle$ should be constant for both the variables; this will be so provided that the random permutation is confined to a narrow region of $\sin\theta$ in reciprocal space. Thus, the variables can be written as $t=I_1/I_2$, $r=I_1/I_1+I_2$, $v=|F_1|/|F_2|$ and $u=|F_1|/(I_1+I_2)^{\frac{1}{2}}$.

* Following Srinivasan & Chandrasekharan (1966), C_A may be called the amplitude correlation.

In fact, this has actually been assumed in expressing (5) and (6) which involve the intensities and amplitudes.

This procedure simplifies the practical steps considerably. The only slight disadvantage here is that the randomization of the pair may be more difficult to achieve if the number of reflexions in any narrow region is not large.

The results discussed earlier were tested with the data of two actual cases, namely the structures of L-ephedrine hydrochloride (Phillips, 1954) and α-rhamnose monohydrate (McGeachin & Beevers, 1957). These two compounds belong to the space group $P2_1$. In both cases, the centric projection about the unique b axis and one of the acentric projections were used. The function $P(t)$ was not tried since the difference in the nature of the curves ${}_CP(t)$ and ${}_AP(t)$ is not prominent. The tests for the other cases are shown in Figs. 2 to 4. The agreement may be seen to be fairly good in all cases.

The values of the linear correlation coefficient of the normalized structure amplitudes y_1 and y_2 were also evaluated and are given in Table 1. In spite of the fact that the values are not very close to zero, the results for the pair are fairly satisfactory as might be seen from the tests described above. It might be mentioned in this connexion that this correlation check was tried actually after the completion of the tests described

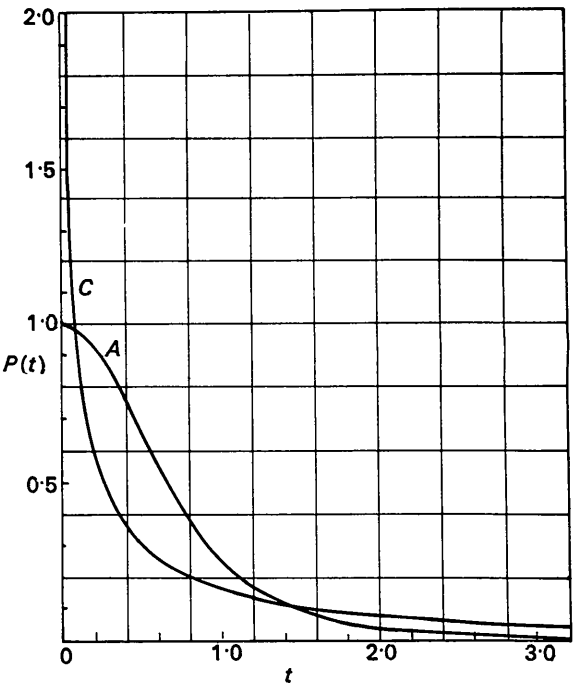

Fig. 1. The probability distribution function $P(t)$. C and A refer to centrosymmetric (centric) and non-centrosymmetric (acentric) cases.

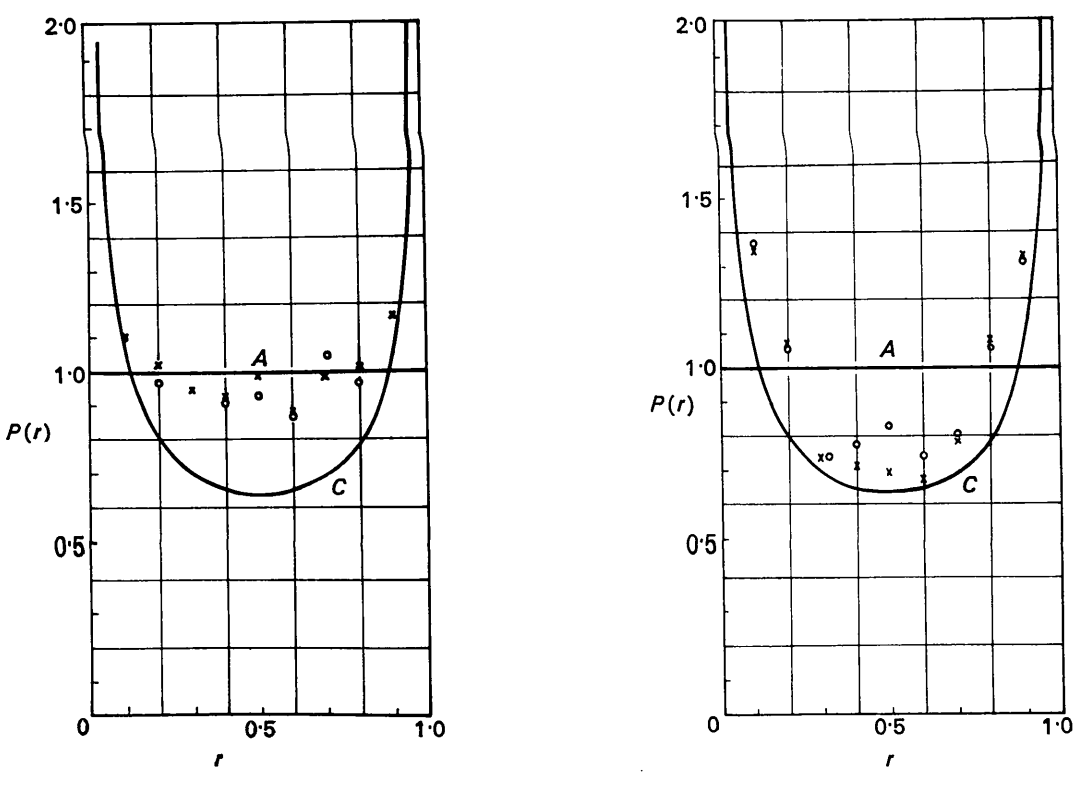

Fig. 2. The distributions $P(r)$ observed in L-ephedrine hydrochloride (crosses) and α-rhamnose monohydrate (closed circles) for (*a*) non-centrosymmetric and (*b*) centrosymmetric cases.

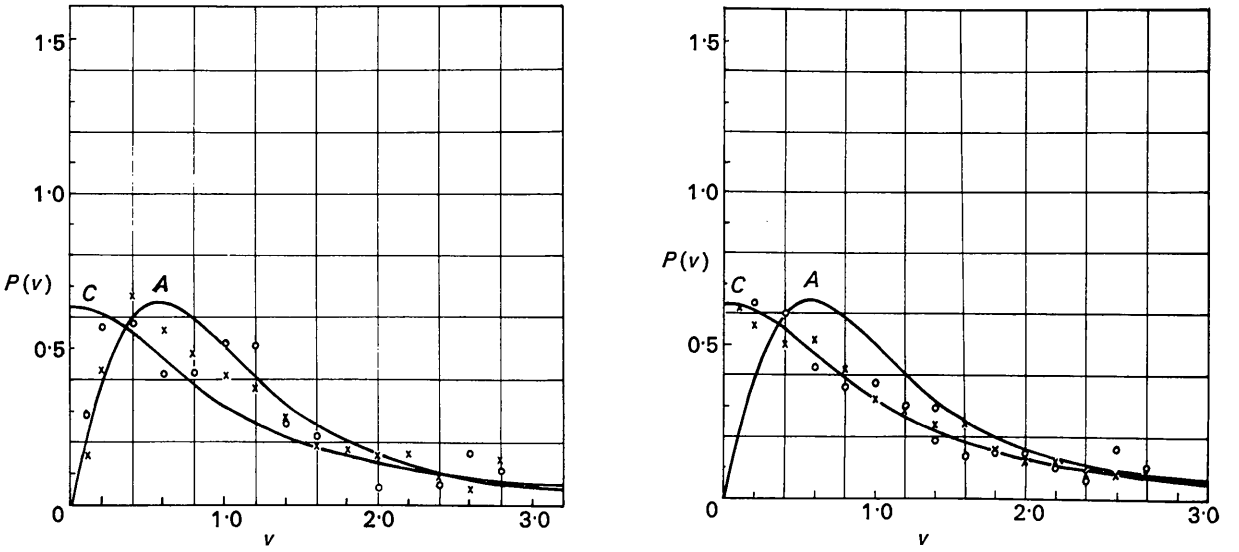

Fig. 3. The distributions $P(v)$ observed in L-ephedrine hydrochloride (crosses) and α-rhamnose monohydrate (closed circles) for (*a*) non-centrosymmetric and (*b*) centrosymmetric cases.

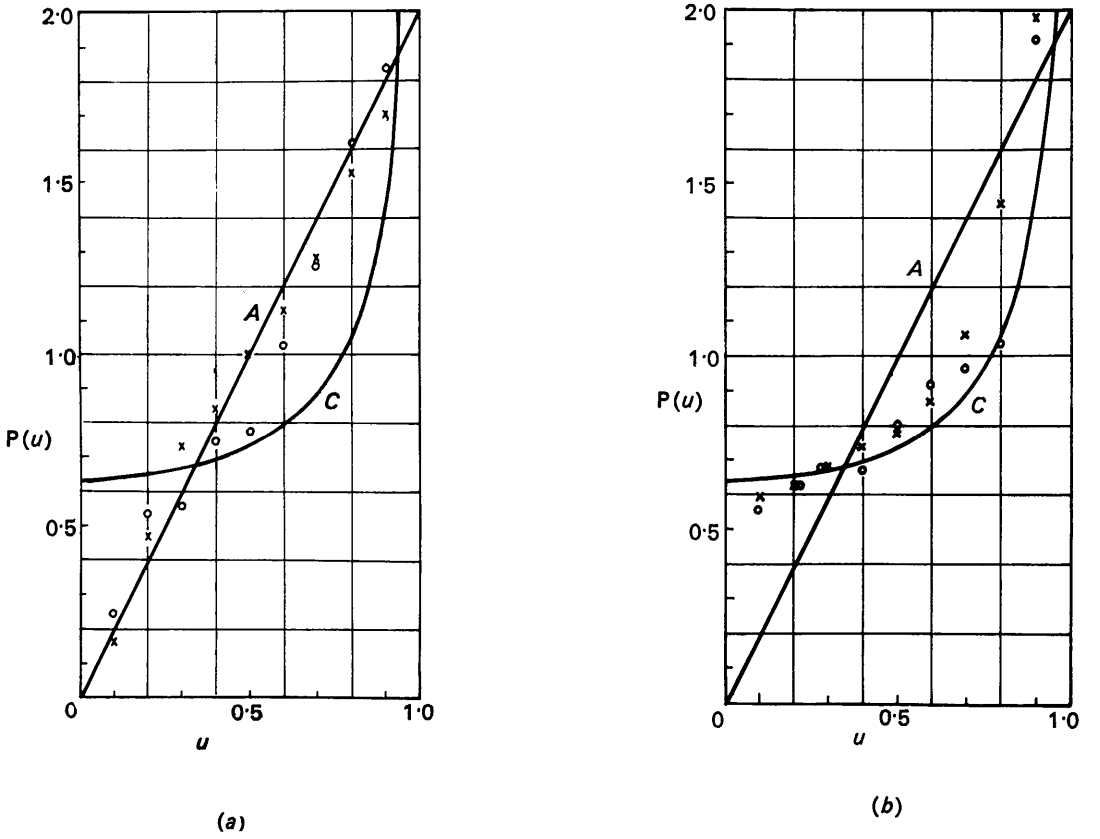

Fig. 4. The distributions $P(u)$ observed in L-ephedrine hydrochloride (crosses) and α-rhamnose monohydrate (closed circles), for (*a*) non-centrosymmetric and (*b*) centrosymmetric cases.

above and also the randomization was done within regions of sin θ, without any use of tables of random numbers. In practice it would be advisable to test the correlation first and make sure that the value of k is as close to zero as possible, which can be expected to lead to better results. The agreement between C_A and ϱ is also seen in Table 1 to be fairly satisfactory.

Table 1. *Observed values of k, C_A and ϱ*

Crystal	k	C_A*	ϱ†
Ephedrine HCl			
Non-centric	0·209	0·691	0·628
Centric	0·240	0·677	0·550
α-Rhamnose monohydrate:			
Non-centric	0·220	0·783	0·715
Centric	0·165	0·610	0·629

* Theoretically $C_A = \varrho$ when $k = 0$.
† Theoretical value of ϱ is 0·637 for centric and 0·785 for non-centric cases.

The theoretical values of R' and R'' for the two cases are given in Table 2, which also contains the observed values with the test data. The agreement is satisfactory except in one case, namely the acentric projection of ephedrine hydrochloride. The observed value (0·538) for this case is rather closer to the centric than to the acentric one. This may be attributed to the statistical fluctuation and the rather imperfect randomization.

Lastly, it appears that the functions R' and R'' may also find useful application in crystal structure refinement. For instance, one may calculate their values between the observed and calculated structure factors (or intensities). If the proposed model is completely wrong and has no relation to the correct structure, the values of R' and R'' are to be expected to be the theoretical ones calculated above. Use of this function for regular structure refinement seems to be an interesting possibility. These aspects are under detailed study and will be reported later.

Table 2. *Theoretical and observed values of the functions R′ and R″*

	R′		R″	
	Centric	Acentric	Centric	Acentric
Theoretical	0·559	0·429	0·727	0·614
L-Ephedrine hydrochloride	0·596	0·484	0·699	0·610
α-Rhamnose monohydrate	0·561	0·538	0·708	0·660

One of us (T.S.) wishes to thank the University of Madras for the award of a Scholarship.

References

KENNEY, J. F. & KEEPING, E. S. (1954). *Mathematics of Statistics*, Part I. New York: Van Nostrand.
MCGEACHIN, H. MCD. & BEEVERS, C. A. (1957). *Acta Cryst.* **10**, 227.
PHILLIPS, D. C. (1954). *Acta Cryst.* **7**, 159.
SRINIVASAN, R. & CHANDRASEKHARAN, R. (1966). *Indian J. Pure Appl. Phys.* **4**, 178.
SRINIVASAN, R. & SUBRAMANIAN, E. (1964). *Acta Cryst.* **17**, 67.
SRINIVASAN, R., SUBRAMANIAN, E. & RAMACHANDRAN, G.N. (1964). *Acta Cryst.* **17**, 1010.
WEATHERBURN, C. E. (1961). *A First Course in Mathematical Statistics*. London: Cambridge Univ. Press.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Acta Cryst. (1966). **21**, 652

The Crystal and Molecular Structure of Tricarbonyltetrakis(trifluoromethyl)cyclopentadienone-iron

BY N. A. BAILEY AND R. MASON
Department of Chemistry, The University, Sheffield, England

(*Received* 24 *February* 1966)

Tricarbonyltetrakis(trifluoromethyl)cyclopentadienone-iron crystallizes in the monoclinic space group $P2_1/c$ with unit-cell dimensions $a = 9{\cdot}300$, $b = 11{\cdot}696$, $c = 16{\cdot}537$ Å, $\beta = 119{\cdot}3°$, $Z = 4$. A three-dimensional Fourier synthesis and least-squares analysis of 968 independent reflexions has reduced the discrepancy index to 0·083. The non-planarity of the substituted cyclopentadienone ligand shows that in a simple valence-bond description of the molecular structure, the formation of localized σ and π bonds between the metal ion and cyclic ligand contributes substantially to the bonding, a conclusion supported by the observed carbon–carbon bond lengths (average e.s.d. 0·034 Å) and bond angles (mean e.s.d. 2°). The three carbonyl groups are not arranged with strict C_{3v} symmetry and a comparison is made with the structures of similar transition metal complexes and interpreted in the light of molecular-orbital theory. Atomic and molecular vibrations and the arrangement of the molecules in the crystal are also discussed.

Introduction

The preparation and characterization of a number of fluorocarbon complexes of transition metal ions have been reported by Boston, Sharp & Wilkinson (1960, 1962), Boston, Grim & Wilkinson (1963) and Dickson & Wilkinson (1964). The reaction of hexafluorobut-2-yne with pentacarbonyliron (Boston *et al.*, 1962) yields a single product which, on the basis of infrared studies, may be formulated as tricarbonyltetrakis(tri-