Acta Cryst. (1965). **18**, 1022

Probability Distribution of the Phases in a Crystal with Heavy Atoms
I. Centrosymmetric Crystal: Probability Distribution of the Sign

BY S. PARTHASARATHY

Department of Physics, University of Madras, Madras 25, *India*

(*Received* 24 *March 1964 and in revised form* 11 *September* 1964)

The probability distribution of the signs of the structure factors of a centrosymmetric crystal containing heavy atoms, in relation to the sign of the contribution from the heavy atoms, is worked out. Four different cases are considered, namely when the number of heavy atoms in the unit cell is 1, 2, 4 and many. A table and curves are given showing the fractional number of reflexions for which the sign is the same as that of the heavy atoms alone, for different contributions from the heavy atoms, for each of the four cases. The theoretical results have also been verified with the data available for a few crystals.

Introduction

The phase problem in crystallography has been attacked by means of various techniques, of which the heavy-atom method, first used by Robertson (1935, 1936) is the earliest and the simplest. The principle of this method is that when the contribution from the heavy atoms dominates the structure amplitude of a crystal, the phase α_N of the contribution from all the N atoms will not be much different from the phase α_P due to the contribution from the P heavy atoms. Thus, the quantity $|F_N| \exp (i\alpha_P)$ may be used as the Fourier coefficient for the first electron-density map. In this connection it will be worth while to study the probability distribution of the deviations of α_P from α_N, that is of $\alpha_N - \alpha_P (= \theta$ say), depending on the heavy-atom contribution and the number P of heavy atoms in the unit cell. The problem becomes particularly interesting in connection with the determination of phases by the anomalous-dispersion method in which the phase α_N of the reflexion is determined (see Ramachandran & Raman, 1956) except for a twofold ambiguity, namely $\alpha_P + \theta$ and $\alpha_P + \pi - \theta$. One method adopted to resolve this ambiguity is to choose the phase angle closer to α_P, that is, to choose between the two possible values of θ that one which is acute (Ramachandran & Raman, 1956). This has been adopted for solving the structures of L(+)-lysine monohydrochloride dihydrate (Raman, 1959) and Factor V 1*a* (Dale *et al.*, 1963). Here again it would be useful to know quantitatively how θ is distributed, in particular the fractional number of reflexions for which θ is less than a specific value say, 90°, for a given heavy-atom contribution.

The theory for the probability distribution of θ for a specific case, namely for a crystal of space group $P1$ with one heavy atom besides a large number (Q) of light atoms, has been worked out by Sim (1957*b*). However, in the present study, the theory is worked out from a more general point of view so that the theory for the other three important cases, namely, for a unit cell containing two heavy atoms and many heavy atoms with centric and acentric configurations, which have not been considered earlier, are also fully worked out. The theoretical results are also used to compare the different cases on a quantitative basis.

The corresponding problem of the centrosymmetric crystal is the probability distribution of the product of the signs (denoted by s) of the total structure factor F_N and that of the heavy atom contribution F_P. This problem for the case of a crystal of space group $P\bar{1}$ with two and many heavy atoms in the unit cell has been worked out by Sim (1957*a*). However, his final expressions are in the form of integrals. In this paper the distributions have been obtained in terms of explicit functions. Further, one more case of common occurrence is considered, namely that of a centrosymmetric crystal containing four heavy atoms in the unit cell. Also a unified method of approach is adopted both for the distribution of θ and of s.

In part I the theory for the distribution of s in a centrosymmetric crystal is worked out, and in part II the corresponding problem in the non-centrosymmetric case, namely the distribution of $\theta = \alpha_N - \alpha_P$, is considered. From the theory it will become clear that the probability functions* in either case depend not only on the contribution from the heavy atoms but also on the number of heavy atoms in the unit cell. Therefore, the different cases dealt with have been named according to the number of heavy atoms in the unit cell with the prefix centric or acentric, depending on whether the crystal is centrosymmetric or non-centrosymmetric respectively.

We consider a centrosymmetric crystal containing P heavy atoms and Q light atoms in the unit cell so that

---

* Since we shall be dealing with both continuous and discrete cases in this paper, we use the term *probability density function* in the case of a continuous variable and the term *probability function* in the discrete case. This follows standard usage (see for example Cramer, 1945). The integral of the probability density function is called the *cumulative function* or *distribution function*.

the total number of atoms in the unit cell is $P+Q=N$. The structure factor of a reflexion $H(\equiv hkl)$ can be written in terms of the contributions from the P and the Q atoms as

$$F_N(H)=F_P(H)+F_Q(H)\,. \qquad (1)$$

The following symbols used in previous publications from this laboratory (Srinivasan, Sarma & Ramachandran, 1963) are employed in this paper as well and may be defined here:

$$\sigma_P^2=\sum_{i=1}^{P} f_{Pi}^2\,,\quad \sigma_Q^2=\sum_{j=1}^{Q} f_{Qj}^2\,,\quad \sigma_N^2=\sigma_P^2+\sigma_Q^2$$

$$\sigma_1^2=\sigma_P^2/\sigma_N^2\,,\quad \sigma_2^2=\sigma_Q^2/\sigma_N^2\,,\quad \sigma_1^2+\sigma_2^2=1\,. \qquad (2)$$

Let s_N and s_P represent the sign parts of F_N and F_P respectively and let s denote the product of the signs s_N and s_P, that is, $s=s_Ns_P$. It is evident that

$$s_P^2=s_N^2=s^2=1 \quad \text{and} \qquad (3)$$

$$s=\pm 1\,, \qquad (4)$$

according as F_N and F_P are of the same or opposite signs respectively. Multiplying (1) by s_P one obtains, using (3),

$$s|F_N|=|F_P|+s_PF_Q\,. \qquad (5)$$

The probability density function of F_Q is known to be

$$P(F_Q)=(1/\sqrt{2\pi\sigma_Q{}^2})\exp(-F_Q^2/2\sigma_Q^2)\,, \qquad (6)$$

from which the probability function of s_PF_Q is given by

$$P(s_PF_Q)=(1/\sqrt{2\pi\sigma_Q{}^2})\exp(-F_Q^2/2\sigma_Q^2)\,, \qquad (7)$$

since $s_P^2=1$. With the use of (7) and (5), the conditional probability function of $s|F_N|$ for a given $|F_P|$ is given by

$$P(s|F_N|;|F_P|)=(1/\sqrt{2\pi\sigma_Q{}^2})\exp[-(s|F_N|-|F_P|)^2/2\sigma_Q^2]\,. \qquad (8)$$

Since the centrosymmetric case is fundamentally a one-dimensional random-walk problem, the probability function of $s|F_N|$ can be interpreted as the joint probability of s and $|F_N|$. However, s is a discrete variable capable of taking only the values ± 1. The joint probability function of s and $|F_N|$ for a given $|F_P|$ is therefore given by

$$P(|F_N|,s;|F_P|)=(1/\sqrt{2\pi\sigma_Q^2})\exp[-(s|F_N|-|F_P|)^2/2\sigma_Q^2]\,. \qquad (9)$$

From (9) the probability function of s for a given $|F_P|$ is given by

$$P(s;|F_P|)=\int_0^\infty P(|F_N|,s;|F_P|)d|F_N|$$

$$=(1/\sqrt{2\pi\sigma_Q^2})\int_0^\infty \exp[-(s|F_N|-|F_P|)^2/2\sigma_Q^2]d|F_N|\,. \qquad (10)$$

If the substitution $y=(s|F_N|-|F_P|)/\sqrt{2}\sigma_Q$ is made in (10) and the integrations are carried out, it becomes

$$P(s;|F_P|)=\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)\,, \qquad (11)$$

where the property that the error function is an odd function, *i.e.* $\operatorname{erf}(sx)=s\operatorname{erf}(x)$, is used. From (11), the probability function of s (independent of $|F_P|$) is given by

$$P(s)=\int_{|F_P|} P(s;|F_P|)P(|F_P|)d|F_P|\,. \qquad (12)$$

It is evident from (12) that the probability function of s depends on the density function of $|F_P|$, which in turn depends on the number of heavy atoms in the unit cell. Hence, four important cases arise in the centrosymmetric space group, namely a crystal containing 1, 2, 4 and many heavy atoms (that is $P=1,2,4$ and many) in addition to a large number of light atoms.

It is obvious that $P(+1)$ and $P(-1)$ are the probabilities that the sign of F_N may be positive or negative with reference to F_P. For simplicity these are denoted by the symbols $P(+)$ and $P(-)$ respectively.

Centric one-atom case

In this case the heavy atom is at the centre of inversion, which may be chosen as the origin. Since $\sigma_P^2=f_P^2$ in this case, the density function of $|F_P|$ is given by

$$P(|F_P|)=\delta(|F_P|-f_P)=\delta(|F_P|-\sigma_P)\,. \qquad (13)$$

Using (13) and (11) in (12), the probability function of s is given by

$$P(s)=\int_0^\infty[\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)]\delta(|F_P|-\sigma_P)d|F_P|$$

$$=\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(\sigma_P/\sqrt{}\sigma_Q)$$

$$=\tfrac{1}{2}+(s/2)\operatorname{erf}(\sigma_1/\sqrt{2}\sigma_2)\,. \qquad (14)$$

Centric two-atom case

In this case the density function of $|F_P|$ is given by

$$P(|F_P|)=(2/\pi\sqrt{2}\sigma_P)/\sqrt{1-(F_P^2/2\sigma_P^2)}\,;\quad 0\le F_P\le\sqrt{2}\sigma_P\,. \qquad (15)$$

Using (15) and (11) in (12), the probability function of s is given by

$$P(s)=\int_0^{\sqrt{2}\sigma_p}[\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)]\frac{2d|F_P|}{\pi\sqrt{2}\sigma_P\sqrt{1-(F_P^2/2\sigma_P^2)}}\,. \qquad (16)$$

If the substitution $x=|F_P|/\sqrt{2}\sigma_P$ is made in (16), it becomes

$$P(s)=\frac{1}{\pi}\int_0^1\frac{dx}{\sqrt{1-x^2}}+\frac{s}{\pi}\int_0^1\frac{\operatorname{erf}(\sigma_1x/\sigma_2)}{\sqrt{1-x^2}}dx\,. \qquad (17)$$

Substituting for $\operatorname{erf}(x)$ in terms of Kummer's confluent hypergeometric series (Sneddon, 1961; p. 46, Problem 11 (iii)], (17) becomes

$$P(s)=\tfrac{1}{2}+\frac{s}{\pi}\int_0^1\frac{2}{\sqrt{\pi}}\frac{\sigma_1}{\sigma_2}x\,{}_1F_1\left(\tfrac{1}{2};\tfrac{3}{2};-\frac{\sigma_1^2x^2}{\sigma_2^2}\right)\frac{dx}{\sqrt{1-x^2}}\,. \tag{18}$$

If the substitution $y=x^2$ is made in (18), it becomes

$$P(s)=\tfrac{1}{2}+\frac{\sigma_1 s}{\pi\sqrt{\pi}\sigma_2}\int_0^1(1-y)^{-\frac{1}{2}}\,{}_1F_1(\tfrac{1}{2};\tfrac{3}{2};-\sigma_1^2y/\sigma_2^2)dy\,,$$

which on integration gives (Sneddon, 1961; p. 47, Problem 16 (i)]

$$P(s)=\tfrac{1}{2}+\frac{\sigma_1 s}{\pi\sqrt{\pi}\sigma_2}\beta(1,\tfrac{1}{2})\,{}_2F_2(\tfrac{1}{2},1;\tfrac{3}{2},\tfrac{3}{2};-\sigma_1^2/\sigma_2^2)\,, \tag{19}$$

where β denotes the beta function. If Sim's notation (Sim, 1957*a*) for the ratio of the r.m.s. contributions of the P and Q atoms to the intensity, namely $(\sigma_1/\sigma_2)=r$ is used, (19) becomes

$$\begin{aligned}P(s)&=\tfrac{1}{2}+\frac{rs}{\pi}\sum_{n=0}^{\infty}\frac{(-r^2)^n}{(2n+1)\Gamma(n+\frac{3}{2})}\\&=0{\cdot}5+0{\cdot}3592\,s[r-0{\cdot}2222\,r^3+0{\cdot}0533\,r^5-\ldots]\,,\end{aligned} \tag{20}$$

which is in agreement with the expression derived by Sim (1957*a*) by a different method.

Centric four-atom case:

Let us take the P-group to contain two independent pairs of similar atoms. In this case the density function of $|F_P|$ is given, in terms of Gauss's hypergeometric series, by (see Appendix A)

$$P(|F_P|)=(1/\pi\sigma_P)\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;1-F_P^2/4\sigma_P^2)\,;\quad 0\le F_P\le 2\sigma_P\,. \tag{21}$$

Using (21) and (11) in (12), the probability function of s is given by

$$P(s)=\int_0^{2\sigma_P}\frac{1}{\pi\sigma_P}\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;1-\frac{F_P^2}{4\sigma_P^2}\right)\left[\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)\right]d|F_P|\,, \tag{22}$$

which simplifies to (see Appendix B, equation (B-3)]

$$P(s)=\tfrac{1}{2}+\frac{2s}{\pi^2}\int_0^{\pi/2}K(\sin\alpha)\operatorname{erf}(\sqrt{2}\sigma_1\cos\alpha/\sigma_2)\sin\alpha\,d\alpha\,, \tag{23}$$

where $K(\sin\alpha)=\int_0^{\pi/2}\frac{d\varphi}{\sqrt{1-\sin^2\alpha\sin^2\varphi}}$ represents the complete elliptic integral of the first kind. The integral in the right-hand side of (23) has been evaluated by a numerical integration method for different values of σ_1^2 and the results are given in the form of graphs (see Fig. 1).

Centric many-atom case

In this case the density function of $|F_P|$ is known to be

$$P(|F_P|)=(2/\sqrt{2\pi\sigma_P^2})\exp(-F_P^2/2\sigma_P^2)\,;\quad 0\le|F_P|\le\infty\,. \tag{24}$$

Using (24) and (11) in (12), the probability function of s is given by

$$P(s)=\int_0^{\infty}[\tfrac{1}{2}+\frac{s}{2}\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)](2/\sqrt{2\pi\sigma_P^2})\exp(-F_P^2/2\sigma_P^2)d|F_P|\,,$$

which on making the substitution $x=|F_P|/\sqrt{2}\sigma_P$ becomes

$$P(s)=\frac{1}{\sqrt{\pi}}\int_0^{\infty}[1+s\operatorname{erf}(\sigma_1x/\sigma_2)]\exp(-x^2)dx\,. \tag{25}$$

If the hypergeometric series expansion of erf(x) is used as before, (25) becomes

$$P(s)=\tfrac{1}{2}+\frac{s}{\sqrt{\pi}}\int_0^{\infty}\frac{2}{\sqrt{\pi}}\frac{\sigma_1}{\sigma_2}x\,{}_1F_1(\tfrac{1}{2};\tfrac{3}{2};-\sigma_1^2x^2/\sigma_2^2)\exp(-x^2)dx\,,$$

which, on substituting $y=x^2$, becomes

$$P(s)=\tfrac{1}{2}+\frac{s}{\pi}\frac{\sigma_1}{\sigma_2}\int_0^{\infty}{}_1F_1(\tfrac{1}{2};\tfrac{3}{2};-\sigma_1^2y/\sigma_2^2)\exp(-y)dy\,. \tag{26}$$

Equation (26) simplifies to

$$P(s)=\tfrac{1}{2}+\frac{s}{\pi}\frac{\sigma_1}{\sigma_2}\,{}_2F_1(\tfrac{1}{2},1;\tfrac{3}{2};-\sigma_1^2/\sigma_2^2)\,. \tag{27}$$

Since ${}_2F_1(\frac{1}{2},1;\frac{3}{2};-x^2)=(\tan^{-1}x)/x$ (Sneddon, 1961, p.42, equation 1 (vii)], (27) becomes

$$P(s)=\tfrac{1}{2}+\frac{s}{\pi}\tan^{-1}(\sigma_1/\sigma_2)\,. \tag{28}$$

It is clear that the probability functions (14), (19), (23) and (28) are in the normalized form since they satisfy the normalization condition $P(+)+P(-)=1$. Further, they must satisfy two other physical conditions, namely that $P(+)=\frac{1}{2}$ when there is no heavy atom in the unit cell and that $P(+)=1$ when the light-atom contribution tends to zero. This means that the following limits should hold:

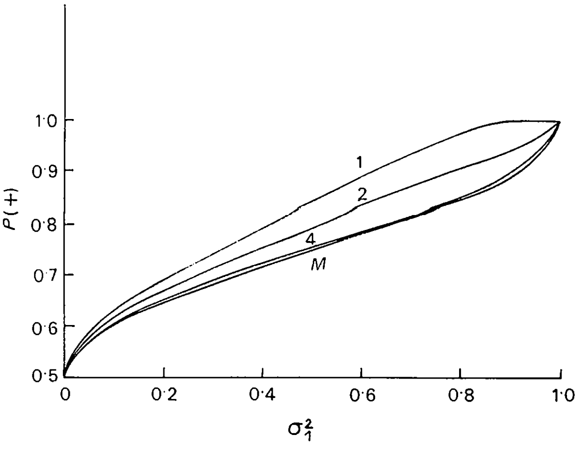

Fig. 1. The $P(+)$ function for the various cases, namely (1) one-atom, (2) two-atom, (4) four-atom and (M) many-atom cases respectively, at different heavy-atom contributions σ_1^2.

$$P(+)\rightarrow\tfrac{1}{2}\quad\text{as}\quad\sigma_1^2\rightarrow 0\quad\text{and}\quad\sigma_2^2\rightarrow 1\,, \tag{29}$$

$$\text{and}\quad P(+)\rightarrow 1\quad\text{as}\quad\sigma_1^2\rightarrow 1\quad\text{and}\quad\sigma_2^2\rightarrow 0\,. \tag{30}$$

It is evident that (29) is satisfied by all the probability functions. Further, it is clear that (30) is satisfied by (14) and (28), while it is not self-evident in the other two cases. They can however be shown to satisfy (30) [see Appendix C, equations (*C*-1) and (*C*-5)].

Discussion of the theoretical results

The probability $P(+)$ has been evaluated as a function of σ_1^2 in all four cases and is tabulated in Table 1. The probabilities are also plotted in Fig. 1 from which it will be seen that all four curves are nearly linear in the middle ranges of σ_1^2, say from $\sigma_1^2=0{\cdot}15$ to $\sigma_1^2=0{\cdot}85$. (The choice of the parameter σ_1^2 instead of $r=\sigma_1/\sigma_2$ (used by Sim; 1957*a* & 1961) is dictated by two reasons: (*a*) that the whole range of possible values can be covered in a graph, and (*b*) that the curves are nearly linear as stated above).

Table 1. *Values of $P(+)$ for the different cases corresponding to different fractional heavy-atom contributions* (σ_1^2)

P	1	2	4	M
$\sigma_1{}^2$				
0·0	0·5	0·5	0·5	0·5
0·2	0·692	0·67	0·651	0·648
0·4	0·793	0·756	0·723	0·718
0·5	0·84	0·795	0·755	0·75
0·6	0·89	0·833	0·785	0·782
0·8	0·977	0·905	0·849	0·852
1·0	1·0	1·0	1·0	1·0

Table 1 and Fig. 1 can be used to determine the fractional number of reflexions for which $s=+1$ if we know the mean value of σ_1^2, that is the contribution to the intensity by the heavy atoms. It is obvious that these data would apply equally well to the case of a light-atom structure in which one knows the positions of P atoms out of the total of N atoms in the unit cell and it is desired to compare the signs of F_P and F_N.

The results of the theory can be used to discuss the general nature of the heavy-atom method from the point of view of the number (P) and the fractional contribution (σ_1^2) of the heavy atoms in the unit cell. In general, since a larger value of $P(+)$ means that more terms are given the correct sign if the sign of F_P is used, a larger value of $P(+)$ would be more favourable for using the heavy-atom method of structure analysis. From Fig. 1 and Table 1 it is easily seen that, for a given value of σ_1^2,

$$P_1(+)>P_2(+)>P_4(+)\simeq P_M(+)$$

where the subscript to the probability function refers to the number of heavy atoms in the P-group. This means that, for a given σ_1^2, the heavy-atom method leads to a better determination of the structure as the number of heavy atoms in the P-group decreases. For example, for $\sigma_1^2=0{\cdot}5$, the value of $P(+)$ is 0·75 in the many-atom case while it is as high as 0·84 in the one-atom case. Incidentally, it is interesting to note that the value in the former case is in agreement with the value given by Lipson & Cochran (1953) purely from physical arguments.

It may be pointed out that a large value of $P(+)$ is not the only criterion for the success of the heavy-atom method. Since the terms for which $|F_P|\simeq 0$ are omitted in the practical computation of the γ'-synthesis the actual distribution of $|F_P|$ is also important in determining the success of the heavy-atom method as this depends on the number of terms that can be safely put in the early Fouriers. The percentage of reflexions for which $y_P(=|F_P|/\sigma_P)$ is less than, say 0·1, is larger for the cases with $P=4$ or many, than the cases with $P=1$ or 2. Thus, from the point of view of including more terms also, the case with a smaller number of heavy atoms in the P-group is more favourable (for a given cut off and a given value of σ_1^2). However, it may be noted that all terms can be included in computing the γ'-synthesis by employing a simple weighting function which becomes useful in improving the resolution of the unknown atoms, especially when the heavy-atom contribution is small (for details see Blow & Crick (1959) and Woolfson (1956)].

From Fig. 1 and Table 1 it is easily seen that for a given number of heavy atoms in the unit cell (*i.e.* P is fixed) the value of $P(+)$ increases as σ_1^2 increases. Thus, the percentage of reflexions whose signs are correctly determined if the sign of F_P is used increases as σ_1^2 increases. Since better resolution of unknown atoms is obtained by using more terms with correct sign, it follows from the above that the resolution of the Q-atoms improves with increasing value of σ_1^2. This result deduced here from qualitative arguments is in agreement with the results obtained by Luzzati (1953) and by Woolfson (1956), who have considered the question of resolution of the unknown atoms in a quantitative way using statistical methods.

It will be seen in Fig. 1 that the curve for the four-atom case is quite close to that of the many-atom case. In order to see how close the distribution functions are for these two cases, they have been plotted in Fig. 2. The distribution for $y=|F_P|/\sigma_P$ in the many-atom case is well-known and is given by

$$P(y)\;=\sqrt{2/\pi}\,\exp\,(-y^2/2)\,. \tag{31}$$

That for the four-atom case has been derived in Appendix I and may be given here. The function is

$$P(y)=\frac{1}{\pi}\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;1-y^2/4)=\frac{2}{\pi}\,K(\sqrt{1-y^2/4}). \tag{32}$$

It will be seen from Fig. 2 that the density function $P(|F_P|)=0$ for $(|F_P|/\sigma_P)>2$ in the four-atom case, while it is finite in this region for the many-atom case. This may possibly explain the peculiar crossing over of the two curves in Fig. 1. It has been pointed out that for $P \geq 6$ the distribution of $y=|F_P|/\sigma_P$ tends to be close to the asymptotic distribution for a large number of atoms (Qurashi, 1953). In view of this it is not necessary to consider these cases separately. In fact from the results shown in Fig. 1, it may be taken that the distribution is close to that of the many-atom case for $P \geq 4$.

The values given in Fig. 1 have also been tested with the experimental data for four crystals, namely, L-tyrosine hydrochloride (Srinivasan, 1959*b*), L-tyrosine hydrobromide (Srinivasan, 1959*a*), α-D-glucosamine hydrobromide (Chandrasekaran & Mallikarjunan, 1963) and α-D-glucosamine hydrochloride (Chandrasekaran & Mallikarjunan, unpublished). All four crystals belong to the space group $P2_1$ (two-atom case) and the $h0l$ reflexions belonging to the centrosymmetric projection were used for the test. For obtaining the mean value of σ_1^2, say $\langle\sigma_1^2\rangle$, the data were divided into various ranges of $(\sin\theta)/\lambda=0$ to 0·1, 0·1 to 0·2, *etc.* and the value of σ_1^2 corresponding to each range was weighted according to the number of reflexions in that range. In this way, the mean value $\langle\sigma_1^2\rangle$ was obtained, which was used for the comparison with theory. The results are shown in Table 2.

Table 2. *Verification of the $P(+)$ function for the centric two-atom case*

Crystal	Space group	$\langle\sigma_1{}^2\rangle$	$P(+)$ Exp.	Theor.
L-tyrosine HBr	$P2_1$	0·84	0·898	0·918
L-tyrosine HCl	$P2_1$	0·47	0·81	0·79
α-D-glucosamine HBr	$P2_1$	0·83	0·879	0·915
α-D-glucosamine HCl	$P2_1$	0·46	0·765	0·78

The author wishes to express his thanks to Professor G. N. Ramachandran and Dr. R. Srinivasan for the useful discussions he had with them during the course of this work. He wishes to thank Dr. K. Venkatesan for having suggested the problem. Thanks are also due to M. Mallikarjunan and R. Chandrasekaran for providing the data on α-D-glucosamine hydrohalides before publication. He also wishes to thank the referees for their useful comments. Finally, his thanks are due to the Council of Scientific and Industrial Research for the award of a Junior Research Fellowship during the tenure of which this work was done.

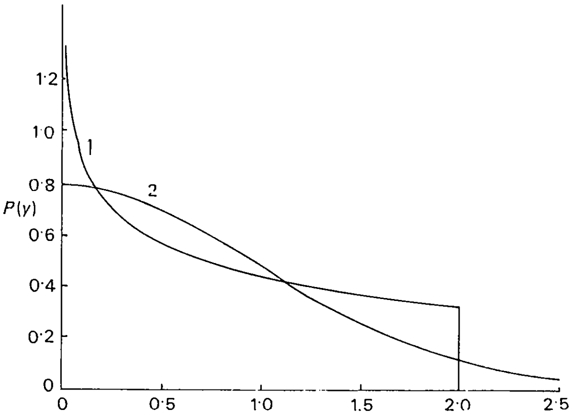

Fig. 2. The probability density function of the normalized variable $y=|F_P|/\sigma_P$ for a unit cell of space group $P\bar{1}$ containing (1) four atoms and (2) many atoms.

APPENDIX A

Distribution of $|F_P|$ for the case of four atoms

The structure factor of a reflexion H for a crystal containing two independent pairs of atoms of the same type in the space group $P\bar{1}$ is given by

$$F_P(H)=2f_P(\cos\theta_1+\cos\theta_2)\,. \qquad (A\text{-}1)$$

Making the substitution $y=F_P/\sigma_P$, (A-1) becomes, since $\sigma_P^2=4f_P^2$,

$$y=\cos\theta_1+\cos\theta_2=x_1+x_2 \text{ (say)};\ |x_i|\leq 1 \text{ and } i=1,2\,. \qquad (A\text{-}2)$$

The density function of x_i is given by

$$P(x_i)=(1/\pi)/\sqrt{1-x_i^2}\,. \qquad (A\text{-}3)$$

Since x_1 and x_2 are independent random variables, the Fourier transform of the density function of y is the product of the Fourier transforms of the density functions of x_1 and x_2. We may denote by $F_e[P(x)]$ and $F_c[P(x)]$ the exponential and cosine Fourier transforms respectively of the function $P(x)$. Then from (A-2) it follows that

$$F_e[P(y)]=F_e[P(x_1)]F_e[P(x_2)]\,. \qquad (A\text{-}4)$$

Since $P(x_i)$ is an even function of x_i, (A-4) can be written

$$F_e[P(y)]=2F_c[P(x_1)]2F_c[P(x_2)]\,. \qquad (A\text{-}5)$$

Now

$$F_c[1/\sqrt{1-x_i^2}]\ =\frac{\pi}{2}J_0(R),$$

(Erdelyi, 1954, p. 11, equation 8) where R is the variable in Fourier space. Thus (A-5) and (A-3) give

$$F_e[P(y)]=[J_0(R)]^2\,. \qquad (A\text{-}6)$$

Taking the inverse Fourier transform, (A-6) gives

$$P(y)=\frac{1}{2\pi}F_e^{-1}[J_0(R)]^2\ =\frac{1}{\pi}F_c^{-1}[J_0(R)]^2, \qquad (A\text{-}7)$$

where $\frac{1}{2\pi}$ is the usual factor occurring in such inverse transformations. Using equation (21), p. 46 of Erdelyi (1954), the density function of y is given by

$$P(y)=\frac{1}{2\pi}P_{-1/2}\left(\frac{y^2}{2}-1\right),\ |y|\leq 2\,. \qquad (A\text{-}8)$$

But

$$P_{-1/2}(z)={}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;\frac{1-z}{2}\right), \qquad (A\text{-}9)$$

(Erdelyi, 1954, p. 370). From (A-9) and (A-8), the density function of y is therefore given by

$$P(y)=\frac{1}{2\pi}{}_2F_1\left((\tfrac{1}{2},\tfrac{1}{2};1;1-\frac{y^2}{4}\right),\ |y|\leq 2\,. \qquad (A\text{-}10)$$

Hence the density function of $|F_P|=\sigma_P|y|$ is given by

$$P(|F_P|)=\frac{1}{\pi\sigma_P}\,{}_2F_1\left((\tfrac{1}{2},\tfrac{1}{2};1;1-\frac{F_P^2}{4\sigma_P^2}\right),$$
$$0\le|F_P|\le 2\sigma_P\,. \quad (A\text{-}11)$$

This function can be shown to be in the normalized form: for

$$\int_0^{2\sigma_P} P(|F_P|)dF_P =$$
$$\int_0^{2\sigma_P}\frac{1}{\pi\sigma_P}\,{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;1-\frac{F_P^2}{4\sigma_P^2}\right)dF_P=I\,(\text{say})\,. \quad (A\text{-}12)$$

On making the substitution $x=1-|F_P|^2/4\sigma_P^2$ in (A-12), it becomes

$$I=\frac{1}{\pi}\int_0^1(1-x)^{-\frac{1}{2}}{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;x)dx\,, \quad (A\text{-}13)$$

which on integration becomes (Sneddon, 1961, p. 47, equation 16(i)]

$$I=\frac{1}{\pi}\,\beta(1,\tfrac{1}{2})\,{}_3F_2(\tfrac{1}{2},\tfrac{1}{2},1;1,\tfrac{3}{2};1)=\frac{2}{\pi}\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};\tfrac{3}{2};1)\,. \quad (A\text{-}14)$$

Since ${}_2F_1(\frac{1}{2},\frac{1}{2};\frac{3}{2};1)=\sin^{-1}1$ (Sneddon, 1961, p. 42, equation 1(vi)], (A-14) becomes

$$I=(2/\pi)\sin^{-1}1=1\,. \quad (A\text{-}15)$$

APPENDIX B

Making use of (A-12) in (22), it becomes

$$P(s)=\tfrac{1}{2}$$
$$+\frac{s}{2\pi\sigma_P}\int_0^{2\sigma_P}{}_2F_1\left(\tfrac{1}{2},\tfrac{1}{2};1;1-\frac{F_P^2}{4\sigma_P^2}\right)\operatorname{erf}(|F_P|/\sqrt{2}\sigma_Q)dF_P$$
$$=\tfrac{1}{2}+\frac{s}{\pi}I_1\ (\text{say})\,. \quad (B\text{-}1)$$

Substituting $\cos\alpha$ for $|F_P|/2\sigma_P$ in I_1, it becomes

$$I_1=\int_0^{\pi/2}{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;\sin^2\alpha)\operatorname{erf}(\sqrt{2}\sigma_1\cos\alpha/\sigma_2)\sin\alpha\,d\alpha$$
$$=\int_0^{\pi/2}\frac{2}{\pi}K(\sin\alpha)\operatorname{erf}(\sqrt{2}\sigma_1\cos\alpha/\sigma_2)\sin\alpha\,d\alpha\,, \quad (B\text{-}2)$$

where $K(\sin\alpha)$ is the complete elliptic integral of the first kind [Sneddon, 1961, p. 42, equation 1(viii)]. Substituting (B-2) in (B-1), it becomes

$$P(s)=\tfrac{1}{2}$$
$$+\frac{2}{\pi^2}s\int_0^{\pi/2}K(\sin\alpha)\operatorname{erf}(\sqrt{2}\sigma_1\cos\alpha/\sigma_2)\sin\alpha\,d\alpha\,. \quad)B\text{-}3)$$

APPENDIX C

To show that (30) is satisfied by the probability function of the two-atom case, one avoids using equation (19) since it is in the form of a power series. If, however, the previous integral (17) is considered, it becomes, when $\sigma_1^2\to 1$,

$$P(+)=\frac{2}{\pi}\int_0^1\frac{dx}{\sqrt{1-x^2}}=1\,. \quad (C\text{-}1)$$

Similarly if the four-atom case is considered, as $\sigma_1^2\to 1$, (23) becomes

$$P(+)=\tfrac{1}{2}+\frac{2}{\pi^2}\int_0^{\pi/2}K(\sin\alpha)\sin\alpha\,d\alpha\,. \quad (C\text{-}2)$$

If the substitution $\sin\alpha=\sqrt{x}$ is made in (C-2), it becomes

$$P(+)=\tfrac{1}{2}+\frac{1}{\pi^2}\int_0^1\frac{K(\sqrt{x})}{\sqrt{1-x}}dx\,. \quad (C\text{-}3)$$

Substituting for $K(\sqrt{x})$ in terms of hypergeometric series as before, (C-3) becomes

$$P(+)=\tfrac{1}{2}+\frac{1}{2\pi}\int_0^1{}_2F_1(\tfrac{1}{2},\tfrac{1}{2};1;x)(1-x)^{-\frac{1}{2}}dx\,, \quad (C\text{-}4)$$

and, making use of (A-13) in (C-4), we get

$$P(+)=\tfrac{1}{2}+\tfrac{1}{2}=1\,. \quad (C\text{-}5)$$

References

Blow, D. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794.

Cramer, H. (1945). In *Mathematical Methods of Statistics*, Princeton Univ. Press.

Dale, D., Hodgkin, D. C. & Venkatesan, K. (1963). In *Crystallography and Crystal Perfection*, Ed: G. N. Ramachandran, p. 237. London: Academic Press.

Erdelyi, A. (Ed.) (1954). In *Tables of Integral Transforms*, Vol. I. New York: McGraw-Hill.

Lipson, H. & Cochran, W. (1953). In *The determination of Crystal Structures*. London: Bell and Sons.

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

Qurashi, M. M. (1953). *Acta Cryst.* **6**, 577.

Ramachandran, G. N. & Raman, S. (1956). *Curr. Sci.*, **25**, 348.

Raman, S. (1959). *Z. Kristallogr.* **111**, 302.

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

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

Sim, G. A. (1957*a*). *Acta Cryst.* **10**, 177.

Sim, G. A. (1957*b*). *Acta Cryst.* **10**, 536.

Sim, G. A. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*. Eds. R. Pepinsky, J. M. Robertson & J. C. Speakman. p. 227. London: Pergamon Press.

Sneddon, I. N. (1961). In *Special Functions of Mathematical Physics and Chemistry*. London: Oliver and Boyd.

Srinivasan, R. (1959*a*). *Proc. Indian Acad. Sci.* A**49**, 340.

Srinivasan, R. (1959*b*). *Proc. Indian Acad. Sci.* A**50**, 19.

Srinivasan, R., Sarma, V. R. & Ramachandran, G. N. (1963). In *Crystallography and Crystal Perfection*. Ed: G. N. Ramachandran, p. 85. London: Academic Press.

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