*Acta Cryst.* (1967). **23**, 18

# Crystal Structure Analysis from a Viewpoint of Information Theory

BY SUKEAKI HOSOYA AND MASAYASU TOKONAMI

*Institute for Solid State Physics, University of Tokyo, Azabu, Minato-ku, Tokyo, Japan*

(*Received* 21 *October* 1966)

Information amount is defined for values of reflexion intensities and Patterson peaks. How the information recovery is made during the procedures of structure analysis is exemplified by a real structure 96*R*-SiC. The characteristic differences between the direct method, especially the statistical method, and the Patterson method are shown. In the usual statistical method the random distribution of atomic positions is assumed. However, limitations such as steric hindrance or molecular forms affect the intensity distribution. Influence due to this complexity is also included to some extent in the present example. From this standpoint, comments are made on several theoretical works on structure analysis.

## Introduction

The information theory was first developed in communication engineering. Although the theory was later successfully applied to light optics, it has so far been used only in a few works in the field of X-ray structure analysis of crystals. The sampling theorem which is fundamental in the information theory was applied by Sayre (1952) to the phase problem. The familiar unit 'bit' devised in the information theory was used by Diamond (1963) for expressing the information amount included in each inequality relation among structure factors. The information amount included in an absolute value of a structure factor was discussed by one of the present authors in his review article on the phase problem (Hosoya, 1964). In the present paper, the notion of the information amount is defined in a more general form, and then applied to the information amount included in peaks of a Patterson function or of a vector set. Some numerical analyses are described for an existing model.

The general idea will be presented using the unitary structure factor $U$. As is well known, the crystal structure factor $F(\boldsymbol{h})$ is expressed by the atomic scattering factor $f_j$ and the atomic coordinate $\mathbf{r}_j$ of a $j$th atom as follows:

$$F(\boldsymbol{h})=\sum_j f_j \exp(2\pi i \boldsymbol{h} \cdot \mathbf{r}_j) .$$

If each $f_j$ is proportional to the average atomic scattering factor $\hat{f}$ as

$$f_j=a_j\hat{f} ,$$

all discussion on $U(\boldsymbol{h})$ is also valid for $F(\boldsymbol{h})$, where $a_j$ is the atomic number of the $j$th atom. It is to be noted that the values of $U$ corresponding to the reciprocal points located too far from the origin of reciprocal space cannot be observed.

Now, two kinds of space, $\mathbf{x}_j$ space and $\mathbf{I}^h$ space, are introduced as follows: $\mathbf{x}_j$ space is defined by a set of $J$ atomic coordinates $(x_1, x_2, \ldots, x_J)$ and $\mathbf{I}^h$ space is defined by a set of $H$ observed values $(I^{h1}, I^{h2}, \ldots, I^{hH})$, where $I^h=|U(\boldsymbol{h})|^2$ will be called the unitary intensity. The crystal structure analysis is eventually the procedure of finding out a point in $\mathbf{x}_j$ space from the point in $\mathbf{I}^h$ space given by the observations.

In the present work, consideration is given as to how the information obtained from the observation of $I^h$ reduced the super-volume, including the solution point, in $\mathbf{x}_j$ space, and how the amount of information can be geometrically defined in a multi-dimensional space. A real structure, 96*R*-SiC, is discussed for which the dimension $J$ of $\mathbf{x}_j$ space is 32, and the dimension $H$ of $\mathbf{I}^h$ space happens to be also 32.

## Definition of information amount

At every stage when some information is given, a super-volume in $\mathbf{x}_j$ space containing points which are possible solutions should shrink to a more limited volume. Then the information amount obtained at each stage may be expressed as $-\log W$ (Shannon, 1948), where $W$ is the ratio of these two super-volumes before and after the shrinkage. When 2 is taken for the base of logarithms, the value of $-\log W$ is expressed in 'bits'.

The information amount can be defined for continuous as well as discontinuous variables. In practical calculations, however, each variable can be dealt with as a quantity quantized in discrete levels. For instance, atomic fractional coordinates are usually expressed by 2–3 decimal digits at an early stage of the analysis and then by 4–5 digits in a refining stage. In other words, the $\mathbf{x}_j$ space can be considered as a set of sampling points forming a lattice. As to the diffraction intensity, the value $I^h$ is also best taken as a quantity quantized in finite degrees of magnitude, especially because the measured values are always more or less affected by errors.

Suppose that the super-volume including the solution point is scanned in $\mathbf{x}_j$ space. Corresponding to the change of $I^h$ due to such a scanning, let $p(I_i^h)$ be the probability that the value $I^h$ has the intensity of the $i$th degree. Once an $\boldsymbol{h}$ reflexion has been known to have the value $I_i^h$, it gives the information amount $-\log$

$[p(I_i^h)]$, and therefore the expected value for the information amount from the $\boldsymbol{h}$ reflexion should be

$$H(\boldsymbol{h}) = -\sum_i p(I_i^h) \log[p(I_i^h)] . \quad (1)$$

Expressions of this type will be used throughout the present work.

To be more general, the information amount given by two reflexions is

$$H(\boldsymbol{h}1,\boldsymbol{h}2) = -\sum_{ij} p(I_i^{h1}, I_j^{h2}) \log[p(I_i^{h1}, I_j^{h2})] , \quad (2)$$

using a joint probability which can be defined analogously to $p(I_i^h)$. Those formula relevant to more than two reflexions are also given in a similar way. The redundancy in the two reflexions is thus defined as

$$R(\boldsymbol{h}1,\boldsymbol{h}2) = H(\boldsymbol{h}1) + H(\boldsymbol{h}2) - H(\boldsymbol{h}1,\boldsymbol{h}2) . \quad (3)$$

The redundancy serves as a measure of the information recovery.

In the same way, the information amount given by a value at a point $\boldsymbol{r}$ of a vector set or a Patterson function can be defined as

$$H(\boldsymbol{r}) = -\sum_i p(V_r^i) \log[p(V_r^i)] , \quad (4)$$

where $V_r^i$ is the value integrated over a small region specified by the position $\boldsymbol{r}$ in the Patterson space, and a suffix $i$ is a degree of magnitude of this peak value.

## Example with 96*R*-SiC

(1) *The estimated number of possible structures*

As is well known, the structure of SiC consists of a stacking of *ABC* layers, being essentially a one-dimensional structure. The structure factor of 96*R*-SiC can be expressed as

$$F(hkl) = f(hkl) \, . \, U(l) ,$$

where $f(hkl)$ is the structure factor for a chemical unit of SiC and $U(l)$ is the unitary structure factor for the one-dimensional crystal:

$$U(l) = \sum_{j=1}^{32} \exp(2\pi i l z_j/96) , \quad (z_j = \text{integer}).$$

Because the approximate value of $|f(hkl)|$ can be easily calculated, $|U(l)|^2$ can be obtained from the observed $|F(hkl)|^2$. It is to be noted here that the maximum value of $|U(l)|$ in this example is normalized, for convenience, to be 32 instead of 1.

The $\mathbf{x}_j$ space for the 96*R*-SiC structure is, therefore, of 32 dimensions, in which only points with coordinates of multiples of 1/96 need to be taken into account.

If there is no limitation at all for the coordinates $z_j$, $32^{96} = 2^{480} \simeq 10^{144}$ kinds of structure may be possible. Actually, there are of course several limitations as follows.

(i) More than one layer cannot occupy an identical position. This limitation is expressed as

$$z_j - z_{j'} \not\equiv 0 \pmod{96} ,$$

which makes possible structures decrease in number down to

$$96!/(32!64!) \simeq 2^{84\cdot6} \simeq 10^{25} .$$

(ii) Because of steric hindrance, a layer *A* cannot follow *A*, and this is the case with *B* and *C*, respectively. This limitation can be expressed as

$$z_j - z_{j'} \not\equiv 1 \pmod{96} ,$$

which makes the possible number of structures decrease to $64!/(32!)^2 \simeq 2^{60\cdot7} \simeq 10^{18}$.

(iii) The present structure has rhombohedral symmetry, and this means that the structure can be specified by an arrangement of *ABC* layers at 32 successive positions. This gives a limitation

$$z_j - z_{j'} \not\equiv 32 \pmod{96} ,$$

which reduces the number of possible arrangements down to $2^{32} \simeq 10^{9\cdot6}$.

(iv) So far only permutation on a line has been considered, but permutation on a circle should further be considered because of the periodic nature of crystals. Moreover, it is usually impossible to distinguish two arrangements on a circle with the reverse order. It is, therefore, sufficient to consider a necklace permutation, which gives us possible structures of about $2^{32}/(96 \times 2) \simeq 2^{24\cdot4} \simeq 10^{7\cdot3} \simeq 22{,}370{,}000$ arrangements or 24·4 bits.

Those structures which have survived the above limitations (i)–(iv) will hereafter be called *the big set*.

(v) We have, in addition, the experimental fact that all polytypes of SiC so far found have structures such that $h$ does not follow $h$ in terms of the Wyckoff–Jagodzinski $h$-$k$ notation, except for the 2*H* wurtzite type (Jagodzinski, 1949; Krishna & Verma, 1965). This limitation gives the condition that $z_{j+1} - z_j \leq 4$ when $z_j$'s are arranged in increasing order. More details about these conditions expressed in $z$'s have been described elsewhere (Tokonami, 1966).

(vi) Among all kinds of necklace arrangements with 96 beads, there are some arrangements with a periodicity of 48 or other shorter lengths. However, these were excluded in advance in an experimental analysis because of a definite periodicity of the crystal being specified from the very beginning. Exactly speaking, this limitation should duly be added at every stage of the limitations (i) to (v). For instance, the possible number of structures is reduced from $2^{32}$ to $2^{32} - 2^{16}$ when the present limitation is added to condition (v). As seen from this example, this modification is always numerically negligible.

All arrangements of $z_j$'s satisfying conditions (i) to (vi) in the above will later be referred to as *the small set*. The number of arrangements in the small set, which have been counted one by one on an electronic computer PC-2 (the commercial name is FACOM 202), is 25780 (corresponding to 14·65 bits).

Summarizing the above, the information amount obtained by condition (i) is $480 - 84\cdot6 \simeq 395$ bits, that by (ii) is $84\cdot6 - 60\cdot7 \simeq 24$ bits, that by (iii) is $60\cdot7 - 32 \simeq 29$

bits, that by (iv) is $32-24{\cdot}4\simeq 8$ and that by (v) is $24{\cdot}4-14{\cdot}65\simeq 10$ bits.

(2) *Preliminary considerations*

The following considerations concern the estimation of the information amount included in the values of $I^l$ and $V_z$, which are used for the present one-dimensional case, corresponding to $I^h$ and $V_r$ respectively.

There are more or less restrictions on the possible values of $I^l$ and $V_z$. Namely, for the former we have $I^{3n}=0$ where $n\not\equiv 0$ (mod 32) and $I^{96n}=1024$. For the latter, being a number of $j-j'$ pairs satisfying $z_j-z_{j'}=z$, we have $V_0=32$, $V_{1/96}=0$ [condition (ii)], and $V_{32/96}=0$ [condition (iii)]. The value of $I^l$ is non-negative and at its maximum 1024, and its non-trivial values are given at $l=1,4,7,\ldots,94$. The values of $V_z$ are integers ranging from 0 to 32 and its non-trivial values are given at $z=2,3,4,\ldots,48$. The values of $I^l$ and $V_z$ at positions other than those shown above are trivial because of either a periodic or a symmetrical character.

(3) *The distribution and information amount of* $I^l$

(*a*) *The case when intensity has been graduated into* 16 *degrees*

The present graduation was adopted so that $I^l$ may be divided with equal probability. This can be done by classifying $I^l$ values into 16 degrees separated at $32\times\ln 16/(16-n)$, where $n=1,2,3,\ldots,15$. Relative abundances for these 16 degrees of $I^l$ obtained with 5000 samples randomly chosen from the big set are listed in Table 1, and those obtained using all members

Table 1. *Frequency distribution of* $I(l)$ *(on the big set)*

| l \ I(l) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 124 | 122 | 107 | 102 | 87 | 83 | 74 | 63 | 50 | 49 | 40 | 33 | 25 | 19 | 11 | 3 |
| 4 | 141 | 136 | 114 | 103 | 90 | 80 | 72 | 59 | 51 | 51 | 39 | 30 | 17 | 10 | 3 | 1 |
| 9 | 155 | 138 | 129 | 110 | 104 | 79 | 73 | 61 | 45 | 35 | 28 | 22 | 10 | 4 | | |
| 10 | 161 | 137 | 128 | 103 | 96 | 89 | 72 | 53 | 51 | 35 | 28 | 19 | 13 | 5 | 3 | |
| 13 | 155 | 136 | 126 | 113 | 100 | 87 | 71 | 55 | 51 | 39 | 27 | 18 | 9 | 4 | 2 | |
| 16 | 67 | 299 | | 194 | | 154 | | 155 | | 49 | 23 | 32 | 10 | 12 | 1 | |
| 19 | 130 | 117 | 113 | 104 | 90 | 92 | 70 | 64 | 58 | 48 | 38 | 32 | 20 | 13 | 5 | |
| 22 | 120 | 112 | 107 | 103 | 94 | 88 | 74 | 68 | 57 | 50 | 40 | 35 | 22 | 16 | 8 | |
| 25 | 114 | 110 | 104 | 86 | 88 | 72 | 68 | 65 | 59 | 54 | 46 | 44 | 33 | 25 | 18 | 7 |
| 28 | 112 | 86 | 98 | 83 | 85 | 71 | 68 | 61 | 62 | 50 | 55 | 48 | 36 | 37 | 27 | 14 |
| 31 | 81 | 81 | 85 | 77 | 78 | 69 | 67 | 64 | 62 | 61 | 50 | 48 | 49 | 46 | 43 | 34 |
| 34 | 83 | 75 | 74 | 63 | 66 | 70 | 61 | 64 | 60 | 62 | 56 | 53 | 49 | 51 | 49 | 60 |
| 37 | 51 | 58 | 49 | 49 | 51 | 52 | 48 | 48 | 54 | 55 | 58 | 62 | 70 | 73 | 89 | 128 |
| 40 | 31 | 34 | 39 | 44 | 47 | 47 | 53 | 23 | 82 | 24 | 75 | 59 | 63 | 80 | 96 | 198 |
| 43 | 22 | 23 | 24 | 24 | 28 | 30 | 36 | 34 | 39 | 39 | 44 | 53 | 60 | 78 | 107 | 353 |
| 46 | 22 | 24 | 20 | 22 | 23 | 28 | 29 | 34 | 32 | 34 | 37 | 44 | 57 | 79 | 101 | 409 |
| 49 | 18 | 16 | 19 | 22 | 25 | 24 | 27 | 29 | 30 | 30 | 40 | 40 | 50 | 64 | 98 | 460 |
| 52 | 19 | 20 | 16 | 23 | 23 | 23 | 24 | 30 | 41 | 33 | 45 | 54 | 47 | 74 | 105 | 417 |
| 55 | 28 | 32 | 32 | 36 | 33 | 40 | 34 | 42 | 44 | 49 | 54 | 65 | 67 | 78 | 102 | 259 |
| 58 | 45 | 41 | 45 | 46 | 46 | 48 | 46 | 51 | 58 | 56 | 65 | 55 | 67 | 72 | 95 | 160 |
| 61 | 61 | 62 | 53 | 56 | 59 | 54 | 57 | 63 | 55 | 58 | 62 | 60 | 60 | 67 | 69 | 98 |
| 64 | 89 | 87 | | 138 | | 124 | 44 | 84 | 44 | 113 | 45 | 47 | 58 | 29 | 43 | 50 |
| 67 | 86 | 86 | 76 | 76 | 77 | 77 | 62 | 68 | 58 | 58 | 57 | 52 | 43 | 48 | 34 | 35 |
| 70 | 115 | 101 | 100 | 94 | 82 | 79 | 72 | 60 | 60 | 48 | 48 | 43 | 34 | 29 | 17 | 10 |
| 73 | 108 | 105 | 99 | 92 | 89 | 83 | 74 | 66 | 56 | 54 | 42 | 41 | 36 | 28 | 14 | 5 |
| 76 | 124 | 114 | 105 | 102 | 96 | 82 | 82 | 66 | 57 | 47 | 44 | 33 | 21 | 14 | 6 | |
| 79 | 140 | 136 | 123 | 104 | 95 | 80 | 74 | 60 | 55 | 42 | 35 | 30 | 14 | 6 | 2 | |
| 82 | 139 | 133 | 109 | 112 | 106 | 80 | 74 | 60 | 56 | 40 | 34 | 26 | 15 | 9 | 2 | |
| 85 | 171 | 142 | 114 | 117 | 89 | 82 | 68 | 62 | 45 | 35 | 27 | 24 | 10 | 6 | 1 | |
| 88 | 171 | 126 | 116 | 112 | 88 | 104 | 59 | 60 | 54 | 31 | 31 | 17 | 13 | 7 | 3 | |
| 91 | 135 | 128 | 125 | 107 | 95 | 84 | 71 | 63 | 47 | 43 | 37 | 31 | 16 | 8 | 4 | 1 |
| 94 | 115 | 113 | 98 | 104 | 90 | 89 | 68 | 68 | 63 | 51 | 44 | 33 | 26 | 20 | 10 | 3 |

Table 2. *Frequency distribution of* $I(l)$ *(on the small set)*

| l \ I(l) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11861 | 6079 | 2839 | 3023 | 1228 | 445 | 138 | 68 | 48 | 47 | 4 | 0 | 0 | 0 | 0 | 0 |
| 4 | 10916 | 6912 | 3841 | 2093 | 1054 | 538 | 282 | 118 | 22 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| 7 | 10472 | 6657 | 4269 | 2283 | 1165 | 622 | 204 | 88 | 19 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 9895 | 6574 | 4117 | 2578 | 1410 | 723 | 324 | 112 | 44 | 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 9208 | 6196 | 3814 | 2743 | 1597 | 921 | 662 | 341 | 155 | 83 | 41 | 10 | 7 | 1 | 1 | 0 |
| 16 | 4299 | 11649 | 0 | 4673 | 0 | 2826 | 0 | 1678 | 0 | 362 | 115 | 135 | 23 | 19 | 1 | 0 |
| 19 | 4674 | 3920 | 3491 | 3048 | 2514 | 2100 | 1768 | 1388 | 1031 | 783 | 505 | 353 | 165 | 36 | 4 | 0 |
| 22 | 2449 | 2426 | 2278 | 2162 | 2041 | 2224 | 1951 | 1786 | 1679 | 1744 | 1542 | 1241 | 1000 | 911 | 342 | 4 |
| 25 | 1196 | 1349 | 1273 | 1217 | 1245 | 1415 | 1435 | 1426 | 1411 | 1545 | 1626 | 1717 | 1817 | 1946 | 2315 | 2857 |
| 28 | 654 | 543 | 625 | 749 | 710 | 728 | 827 | 807 | 1146 | 922 | 1150 | 1394 | 1556 | 2111 | 2952 | 8906 |
| 31 | 801 | 739 | 770 | 827 | 869 | 896 | 977 | 1007 | 1039 | 1137 | 1249 | 1466 | 1714 | 1827 | 2730 | 7732 |
| 34 | 1304 | 1213 | 1067 | 1077 | 1195 | 1131 | 1119 | 1233 | 1188 | 1237 | 1354 | 1476 | 1708 | 1983 | 2401 | 5094 |
| 37 | 1132 | 1091 | 1209 | 1113 | 1238 | 1245 | 1322 | 1377 | 1395 | 1473 | 1550 | 1765 | 1868 | 2072 | 2445 | 3485 |
| 40 | 698 | 2001 | 852 | 424 | 2994 | 263 | 1074 | 10 | 3949 | 28 | 1478 | 2263 | 1214 | 2816 | 1983 | 3733 |
| 43 | 979 | 941 | 1038 | 1064 | 1108 | 1085 | 1121 | 1220 | 1326 | 1373 | 1554 | 1591 | 1854 | 2088 | 2646 | 4792 |
| 46 | 836 | 795 | 820 | 994 | 899 | 943 | 1011 | 1011 | 1164 | 1285 | 1423 | 1558 | 1691 | 2066 | 2841 | 6443 |
| 49 | 836 | 722 | 803 | 878 | 971 | 934 | 993 | 1015 | 1089 | 1230 | 1428 | 1540 | 1700 | 2041 | 2900 | 6700 |
| 52 | 1009 | 787 | 929 | 1030 | 1029 | 1087 | 1069 | 1205 | 1292 | 1349 | 1339 | 1571 | 1883 | 2178 | 2595 | 5428 |
| 55 | 1109 | 1027 | 1107 | 1200 | 1217 | 1183 | 1295 | 1276 | 1397 | 1445 | 1589 | 1725 | 1780 | 2061 | 2565 | 3804 |
| 58 | 1207 | 1035 | 1257 | 1098 | 1419 | 1117 | 1376 | 1426 | 1351 | 1488 | 1509 | 1814 | 1870 | 2099 | 2223 | 3491 |
| 61 | 1109 | 1057 | 1041 | 1063 | 1141 | 1155 | 1208 | 1262 | 1344 | 1443 | 1596 | 1739 | 1842 | 2084 | 2487 | 4209 |
| 64 | 934 | 937 | 0 | 1689 | 0 | 1496 | 797 | 1425 | 806 | 2486 | 1058 | 1059 | 2153 | 1464 | 2717 | 6759 |
| 67 | 605 | 613 | 592 | 671 | 661 | 730 | 774 | 832 | 938 | 951 | 1187 | 1299 | 1515 | 2017 | 2997 | 9398 |
| 70 | 954 | 963 | 965 | 1089 | 1063 | 1079 | 1095 | 1258 | 1320 | 1351 | 1551 | 1557 | 1817 | 2130 | 2647 | 4941 |
| 73 | 1985 | 1878 | 1855 | 1922 | 1866 | 1856 | 1782 | 1801 | 1775 | 1679 | 1658 | 1606 | 1479 | 1281 | 1024 | 333 |
| 76 | 3915 | 3362 | 3126 | 2757 | 2554 | 2232 | 1937 | 1647 | 1295 | 1049 | 821 | 628 | 371 | 86 | 0 | 0 |
| 79 | 5101 | 7408 | 3318 | 2354 | 2000 | 2274 | 1072 | 819 | 537 | 507 | 187 | 130 | 58 | 13 | 2 | 0 |
| 82 | 8628 | 6997 | 2642 | 3030 | 1937 | 683 | 722 | 680 | 145 | 111 | 165 | 13 | 24 | 2 | 1 | 0 |
| 85 | 9605 | 6465 | 4175 | 2561 | 1538 | 813 | 366 | 192 | 46 | 14 | 5 | 0 | 0 | 0 | 0 | 0 |
| 88 | 10782 | 6066 | 4163 | 2523 | 1225 | 749 | 172 | 75 | 23 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |
| 91 | 10998 | 6585 | 3897 | 2216 | 1202 | 548 | 215 | 81 | 29 | 8 | 1 | 0 | 0 | 0 | 0 | 0 |
| 94 | 11105 | 6839 | 4015 | 1791 | 829 | 573 | 449 | 151 | 23 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |

of the small set are listed in Table 2. The information amount $H(l)$ included in $I^l$ for all members of each set calculated by use of formula (1) is listed in Table 3. The simultaneous distribution of $I^{l1}$ and $I^{l2}$ has been calculated for several pairs of $l1$ and $l2$ reflexions; then the information amount $H(l1,l2)$ [formula (2)] and the redundancy $R(l1,l2)$ [formula (3)] have also been calculated. Representative pairs which may be expected to have a heavy correlation are those with $l1:l2=1:2$ and $l2=l1+3$ (neighbouring reflexions). Some other pairs which may not be expected to have even a slight correlation were also chosen for comparison. According to the result shown in Table 4, correlation can hardly be found except for a pair of reflexions with $l1=16$ and $l2=32$ which are two larger available divisors of 96.

The value of $H(l)$ averaged over 32 $l$ values is 3·6 and 3·2 bits respectively for the big set and for the small set. This value multiplied by the number of reflexions is far more than 24·4 and 14·65 bits to be recovered. It means that there should exist much redundancy in a whole set of $I^l$'s. From the fact that redundancy hardly exists between any pair of $I^l$'s except $I^{16}$ and $I^{32}$, a large amount of redundancy should exist mostly among simultaneous distributions of three or more $I^l$'s.

Table 3. *Information amount in the case when the intensity is classified into* 16 *degrees*

| $l$ | $H(l)$ in the big set | $H(l)$ in the small set |
|---|---|---|
| 1 | 3·70 | 2·13 |
| 4 | 3·58 | 2·16 |
| 7 | 3·46 | 2·19 |
| 10 | 3·48 | 2·29 |
| 13 | 3·46 | 2·51 |
| 16 | 2·74 | 2·18 |
| 19 | 3·63 | 3·33 |
| 22 | 3·67 | 3·81 |
| 25 | 3·79 | 3·95 |
| 28 | 3·87 | 3·36 |
| 31 | 3·96 | 3·53 |
| 34 | 3·98 | 3·82 |
| 37 | 3·94 | 3·91 |
| 40 | 3·76 | 3·53 |
| 43 | 3·35 | 3·82 |
| 46 | 3·17 | 3·66 |
| 49 | 3·00 | 3·63 |
| 52 | 3·12 | 3·76 |
| 55 | 3·64 | 3·89 |
| 58 | 3·88 | 3·92 |
| 61 | 3·98 | 3·86 |
| 64 | 3·66 | 3·47 |
| 67 | 3·95 | 3·30 |
| 70 | 3·81 | 3·80 |
| 73 | 3·79 | 3·94 |
| 76 | 3·66 | 3·49 |
| 79 | 3·54 | 2·96 |
| 82 | 3·55 | 2·58 |
| 85 | 3·46 | 2·35 |
| 88 | 3·48 | 2·21 |
| 91 | 3·58 | 2·17 |
| 94 | 3·72 | 2·15 |
| Mean | 3·6 | 3·2 |

Table 4. *Information amount of simultaneous distribution of* $I(l)$

The values of $l,l'$ such as $2,5,8,\ldots,l=(3n-1),\ldots$ correspond to $94,91,88,\ldots,(96-l),\ldots$ in other tables, respectively.

| | | On the big set | | On the small set | |
|---|---|---|---|---|---|
| $l$ | $l'$ | $H(l,l')$ | $H(l)+H(l')-H(l,l')$ | $H(l,l')$ | $H(l)+H(l')-H(l,l')$ |
| 1 | 2 | 7·36 | 0·05 | | |
| 2 | 4 | 7·26 | 0·05 | 4·25 | 0·06 |
| 4 | 8 | 7·02 | 0·04 | 4·36 | 0·01 |
| 8 | 16 | 6·15 | 0·06 | | |
| 16 | 32 | 5·81 | 0·58 | 5·27 | 0·38 |
| 5 | 10 | 7·02 | 0·04 | | |
| 10 | 20 | 7·10 | 0·05 | | |
| 20 | 40 | 7·37 | 0·04 | | |
| 40 | 80 | 6·34 | 0·15 | | |
| 7 | 14 | 6·97 | 0·03 | | |
| 14 | 28 | 7·38 | 0·04 | | |
| 19 | 41 | 7·22 | 0·05 | 7·20 | 0·02 |
| 34 | 37 | 7·88 | 0·05 | 7·70 | 0·03 |

*(b) The case when intensity has been divided into only two degrees, strong and weak*

In the previous section $H(l)$ was found to be 3·6 bits on average for the big set. Similar calculations gave $H(l)=2·7$, 1·7 and 0·8 bits when the number of degrees of intensity was reduced to 8, 4 and 2 respectively. As is to be expected, less redundancy is obtained when the intensity is specified with less accuracy. In other words, the better the accuracy of measurements, the easier the structure analysis because of the more redundancy.

If the intensity is specified with less and less accuracy, the total information given by $I^l$'s for all $l$ finally does not reach the necessary information amount required to solve the structure or at least to determine the structure uniquely apart from homometric ones. In the present case where the intensity is specified by two degrees, a sum of $H(l)$ on $l$ is 25·34 and 18·04 bits for the big set and the small set respectively. These figures are probably less than those figures, 24·5 and 14·65, to be recovered if redundancy is subtracted. To speak in terms of averages, the structure cannot be determined in this case.

Actually, however, it is possible to determine the structure even in such a case, provided that the intensity distribution of reflexions has marked characteristics as seen in the example of 96*R*-SiC. This real structure (Tokonami, 1966) consists mostly of the 6*H* type with some 21*R*. The observed data classified into strong and weak show more or less disagreement with $I^l$ values calculated for every member of the small set. The number of reflexions which showed such a discrepancy was counted by checking each member of the small set. The statistics of all members as regards the number of reflexions which showed a discrepancy is listed in Table 5. Fortunately only one structure was found to show no discrepancy, while the three next nearly similar structures already showed disagreement for two reflexions. The majority of members showed discrepancies for 10 to 14 reflexions. Although no survey was

carried out for the big set, it seems probable that discrepancies may occur more frequently and that still only one member may show a complete coincidence.

(4) *A distribution of $V_z$ and its information amount*

Using about 8000 samples randomly chosen from the big set, relative abundances of $V_z$ values have been calculated as shown in Table 6 and the information amount $H(z)$ [formula (4)] has also been calculated as in Table 7. Because the average amount of information for peaks at 48 positions is 3·6, this value multiplied by 48 is much larger than the information amount to be recovered, which is about 25 bits. The simultaneous distribution of $V_{2/96}$ and $V_{3/96}$ at positions near the origin has been calculated as in Table 8. Because $H(2/96, 3/96) = 5{\cdot}22$ bits and $H(2/96) = 2{\cdot}71$ and $H(3/96) = 3{\cdot}64$ in Table 7, the amount of redundancy between these two peaks is found to be 1·13 bits. In the same way, $R(2/96, 4/96) = 0{\cdot}61$ and $R(3/96, 4/96) = 0{\cdot}78$. It may be interesting that these two $V$'s are as large as $-0{\cdot}83$ for a usual correlation coefficient, being different from the case of $I^{l1}$ and $I^{l2}$. Such a big redundancy mainly comes from correlation between peaks at two positions. This fact seems consistent with the success in the systematic method of unravelling a periodic vector set using $V$ values one by one from the one nearest to the origin (Tokonami & Hosoya, 1965). In addition, $H(z)$ values for the small set are shown in Table 9.

Table 5. *Distribution of structures with various numbers of discrepancies (on the small set, with two degrees of intensity)*

| Number of discrepancies | Number of structures |
|---|---|
| 0 | 1 |
| 1 | 0 |
| 2 | 3 |
| 3 | 3 |
| 4 | 13 |
| 5 | 36 |
| 6 | 101 |
| 7 | 320 |
| 8 | 756 |
| 9 | 1811 |
| 10 | 3208 |
| 11 | 4429 |
| 12 | 4839 |
| 13 | 4411 |
| 14 | 3110 |
| 15 | 1627 |
| 16 | 782 |
| 17 | 250 |
| 18 | 67 |
| 19 | 10 |
| 20 | 3 |
| 21—32 | 0 |

Table 7. *The information amount contained in $V_{z/96}$ (on the big set)*

| $z$ | $H(z)$ |
|---|---|
| 1 | 0·00 |
| 2 | 2·71 |
| 3 | 3·64 |
| 4 | 4·01 |
| 5 | 3·92 |
| 6 | 4·05 |
| 7 | 3·97 |

Table 6. *Relative distribution of $V_{z/96}$*

| $z$ \ $V_{z/96}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1000 |
| 1 | 1000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | | | | | | | | | | 9 | | 46 | | 118 | | 212 | | 245 | | 203 | | 110 | | 40 | | 11 | | 1 | | | | | |
| 3 | 9 | 24 | 52 | 81 | 97 | 118 | 130 | 123 | 101 | 84 | 64 | 45 | 32 | 12 | 12 | 7 | | | | | | | | | | | | | | | | | |
| 4 | | | | | | 6 | 12 | 22 | 43 | 54 | 65 | 80 | 89 | 95 | 90 | 85 | 91 | 67 | 64 | 36 | 37 | 19 | 15 | 4 | 8 | | 1 | | | | | | |
| 5 | 2 | 12 | 18 | 35 | 61 | 71 | 83 | 87 | 105 | 102 | 100 | 83 | 68 | 52 | 38 | 33 | 16 | 13 | 5 | 2 | 1 | | | | | | | | | | | | |
| 6 | | | | 3 | 8 | 20 | 35 | 51 | 61 | 76 | 87 | 96 | 96 | 89 | 80 | 66 | 64 | 47 | 42 | 21 | 25 | 9 | 7 | 1 | 3 | | | | | | | | |
| 7 | 1 | 6 | 12 | 23 | 42 | 56 | 73 | 85 | 98 | 96 | 91 | 92 | 81 | 71 | 52 | 38 | 32 | 22 | 11 | 6 | 1 | 1 | | | | | | | | | | | |
| 8 | | | 2 | 4 | 14 | 28 | 45 | 63 | 71 | 86 | 90 | 92 | 91 | 89 | 73 | 62 | 53 | 40 | 34 | 21 | 14 | 4 | 10 | 1 | | | | | | | | | |
| 9 | | 4 | 11 | 17 | 27 | 49 | 66 | 88 | 101 | 106 | 102 | 91 | 78 | 73 | 55 | 45 | 27 | 21 | 16 | 7 | 4 | 1 | | | | | | | | | | | |
| 10 | | 1 | 4 | 7 | 18 | 31 | 41 | 51 | 64 | 92 | 99 | 104 | 97 | 96 | 80 | 63 | 51 | 34 | 26 | 13 | 11 | 2 | 3 | | | | | | | | | | |
| 11 | | 1 | 6 | 14 | 22 | 35 | 64 | 92 | 113 | 111 | 112 | 106 | 83 | 66 | 49 | 36 | 28 | 19 | 13 | 10 | 4 | 3 | 2 | | | | | | | | | | |
| 12 | | 1 | 3 | 11 | 17 | 27 | 44 | 53 | 71 | 83 | 111 | 103 | 104 | 100 | 78 | 67 | 48 | 25 | 22 | 9 | 8 | 2 | | | | | | | | | | | |
| 13 | | | 3 | 11 | 20 | 37 | 55 | 80 | 104 | 118 | 111 | 108 | 92 | 81 | 53 | 38 | 25 | 22 | 13 | 6 | 7 | 1 | 1 | | | | | | | | | | |
| 14 | | | 1 | 4 | 12 | 27 | 40 | 52 | 77 | 107 | 110 | 116 | 108 | 92 | 78 | 60 | 47 | 29 | 18 | 7 | 3 | | 1 | | | | | | | | | | |
| 15 | | | 1 | 6 | 15 | 25 | 51 | 73 | 103 | 115 | 129 | 126 | 112 | 91 | 60 | 42 | 20 | 12 | 6 | 1 | | | | | | | | | | | | | |
| 16 | | | | 1 | 4 | 14 | 28 | 46 | 74 | 103 | 127 | 147 | 145 | 125 | 99 | 50 | 27 | | | | | | | | | | | | | | | | |
| 17 | | | 1 | 3 | 9 | 25 | 48 | 70 | 100 | 122 | 141 | 131 | 114 | 89 | 62 | 38 | 20 | 10 | 4 | 2 | | | | | | | | | | | | | |
| 18 | | | | 4 | 11 | 24 | 38 | 58 | 88 | 110 | 125 | 131 | 110 | 104 | 74 | 50 | 31 | 19 | 8 | 3 | 2 | | | | | | | | | | | | |
| 19 | | 1 | 1 | 3 | 14 | 30 | 50 | 77 | 103 | 112 | 116 | 114 | 103 | 81 | 68 | 46 | 31 | 16 | 13 | 4 | 3 | 1 | | | | | | | | | | | |
| 20 | | 1 | 4 | 10 | 16 | 31 | 44 | 70 | 92 | 100 | 112 | 113 | 98 | 92 | 74 | 50 | 35 | 21 | 13 | 6 | 4 | 1 | 1 | | | | | | | | | | |
| 21 | | | 2 | 7 | 12 | 28 | 51 | 83 | 95 | 111 | 111 | 102 | 96 | 78 | 66 | 47 | 38 | 22 | 17 | 8 | 7 | 3 | 1 | 1 | | | | | | | | | |
| 22 | | 2 | 7 | 11 | 23 | 36 | 50 | 74 | 83 | 99 | 106 | 112 | 106 | 87 | 67 | 50 | 33 | 20 | 11 | 7 | 4 | | | | | | | | | | | | |
| 23 | | | 1 | 7 | 18 | 27 | 43 | 66 | 91 | 106 | 118 | 108 | 112 | 84 | 68 | 47 | 32 | 25 | 15 | 8 | 7 | 1 | 2 | | | | | | | | | | |
| 24 | | 3 | 6 | 10 | 17 | 29 | 49 | 71 | 88 | 114 | 113 | 126 | 113 | 84 | 61 | 41 | 26 | 16 | 11 | 5 | 3 | 1 | | | | | | | | | | | |
| 25 | | 1 | 2 | 5 | 16 | 28 | 40 | 60 | 84 | 101 | 116 | 126 | 115 | 98 | 68 | 46 | 34 | 19 | 14 | 6 | 5 | 1 | 3 | | | | | | | | | | |
| 26 | 1 | 3 | 5 | 11 | 21 | 37 | 58 | 82 | 106 | 120 | 116 | 107 | 89 | 74 | 52 | 38 | 30 | 17 | 11 | 5 | 4 | 1 | | | | | | | | | | | |
| 27 | | | 1 | 5 | 8 | 21 | 36 | 62 | 74 | 90 | 103 | 117 | 113 | 106 | 83 | 66 | 40 | 26 | 18 | 9 | 7 | 1 | 1 | | | | | | | | | | |
| 28 | 1 | 4 | 9 | 20 | 34 | 53 | 82 | 103 | 120 | 122 | 108 | 94 | 77 | 58 | 44 | 22 | 16 | 11 | 7 | 3 | 1 | | | | | | | | | | | | |
| 29 | | | | 1 | 3 | 7 | 16 | 25 | 44 | 75 | 99 | 111 | 118 | 134 | 108 | 83 | 63 | 41 | 30 | 17 | 7 | 5 | | 1 | | | | | | | | | |
| 30 | 2 | 11 | 25 | 57 | 88 | 115 | 132 | 143 | 124 | 96 | 64 | 54 | 39 | 23 | 9 | 5 | 2 | 1 | 1 | | | | | | | | | | | | | | |
| 31 | | | | | | | | | | 11 | | | 137 | | | 462 | | | 325 | | | 61 | | | 1 | | | | | | | | |
| 32 | 1000 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| 33 | | | | | | | | | 1 | | | 61 | | | 325 | | | 462 | | | 137 | | | 11 | | | | | | | | | |
| 34 | 3 | 13 | 32 | 68 | 98 | 134 | 127 | 127 | 107 | 96 | 74 | 43 | 29 | 17 | 14 | 6 | 1 | | | | | | | | | | | | | | | | |
| 35 | | | | 1 | 2 | 6 | 14 | 25 | 37 | 59 | 84 | 99 | 111 | 123 | 105 | 97 | 84 | 59 | 42 | 22 | 13 | 3 | 1 | | | | | | | | | | |
| 36 | | 2 | 12 | 26 | 52 | 72 | 99 | 108 | 122 | 110 | 102 | 80 | 66 | 49 | 35 | 19 | 15 | 9 | 6 | 3 | 1 | 1 | | | | | | | | | | | |
| 37 | | | 2 | 4 | 8 | 16 | 29 | 41 | 64 | 86 | 101 | 114 | 116 | 106 | 84 | 65 | 52 | 42 | 22 | 20 | 8 | 6 | 1 | | | | | | | | | | |
| 38 | | 3 | 8 | 23 | 38 | 51 | 73 | 90 | 106 | 115 | 110 | 98 | 76 | 65 | 50 | 33 | 22 | 11 | 9 | 3 | 3 | | | | | | | | | | | | |
| 39 | | | 2 | 3 | 11 | 23 | 44 | 53 | 74 | 96 | 111 | 104 | 102 | 89 | 74 | 60 | 49 | 37 | 20 | 19 | 7 | 7 | | 1 | | | | | | | | | |
| 40 | | 9 | 10 | 18 | 31 | 42 | 61 | 84 | 93 | 111 | 109 | 96 | 87 | 76 | 60 | 41 | 26 | 14 | 10 | 5 | 3 | | | | | | | | | | | | |
| 41 | | | 2 | 5 | 13 | 31 | 49 | 71 | 87 | 94 | 100 | 101 | 99 | 86 | 67 | 59 | 40 | 31 | 15 | 17 | 5 | 12 | | 4 | | | | | | | | | |
| 42 | 3 | 8 | 13 | 20 | 26 | 41 | 54 | 72 | 89 | 104 | 109 | 101 | 90 | 74 | 61 | 47 | 34 | 19 | 12 | 6 | 4 | 1 | | | | | | | | | | | |
| 43 | 1 | 2 | 7 | 13 | 26 | 50 | 75 | 91 | 99 | 104 | 104 | 84 | 84 | 66 | 54 | 35 | 32 | 17 | 19 | 4 | 12 | 1 | 6 | | 1 | | | | | | | | |
| 44 | 4 | 8 | 15 | 19 | 30 | 41 | 51 | 70 | 91 | 94 | 109 | 99 | 89 | 78 | 69 | 44 | 30 | 20 | 13 | 7 | 4 | 1 | | | | | | | | | | | |
| 45 | | 1 | 5 | 12 | 23 | 36 | 43 | 70 | 77 | 94 | 93 | 101 | 86 | 85 | 69 | 60 | 39 | 32 | 17 | 20 | 5 | 11 | 1 | 5 | | 1 | | | | | | | |
| 46 | 5 | 10 | 17 | 22 | 36 | 54 | 70 | 78 | 88 | 86 | 85 | 83 | 83 | 65 | 57 | 45 | 34 | 22 | 23 | 9 | 10 | 2 | 2 | | | | | | | | | | |
| 47 | 1 | 6 | 10 | 23 | 33 | 44 | 48 | 61 | 68 | 77 | 83 | 81 | 74 | 69 | 63 | 62 | 47 | 44 | 24 | 27 | 9 | 16 | 2 | 11 | | 4 | | 1 | | | | | |
| 48 | 27 | | 50 | | 99 | | 125 | | 145 | | 147 | | 127 | | 103 | | 74 | | 46 | | 28 | | 14 | | 4 | | 1 | | | | | | |

Table 7 (*cont.*)

| $z$ | $H(z)$ |
|---|---|
| 8 | 4·03 |
| 9 | 3·95 |
| 10 | 3·96 |
| 11 | 3·88 |
| 12 | 3·91 |
| 13 | 3·85 |
| 14 | 3·80 |
| 15 | 3·64 |
| 16 | 3·38 |
| 17 | 3·57 |
| 18 | 3·67 |
| 19 | 3·76 |
| 20 | 3·84 |
| 21 | 3·87 |
| 22 | 3·87 |
| 23 | 3·84 |
| 24 | 3·80 |
| 25 | 3·79 |
| 26 | 3·83 |
| 27 | 3·81 |
| 28 | 3·80 |
| 29 | 3·71 |
| 30 | 3·57 |
| 31 | 1·77 |
| 32 | 0·00 |
| 33 | 1·77 |
| 34 | 3·58 |
| 35 | 3·76 |
| 36 | 3·79 |
| 37 | 3·86 |
| 38 | 3·86 |
| 39 | 3·91 |
| 40 | 3·91 |
| 41 | 3·95 |
| 42 | 3·98 |
| 43 | 3·97 |
| 44 | 3·99 |
| 45 | 4·05 |
| 46 | 4·13 |
| 47 | 4·26 |
| 48 | 3·38 |

### General conclusion expected from the above example

As seen in (3*a*) and (4) in the previous section, a set of $I^l$ values has a stronger correlation than a set of $V_z$ values has between their two respective members, while these sets should be equivalent in amount of information regarding the structure. This fact seems to be valid also in values of $I^h$ and $V_r$ of general crystals. In this connection, it should be noted that the sampling points $\boldsymbol{r}$'s for $V_r$ need not be chosen more densely than required by the sampling theorem (Shannon, 1948). In terms of the more general case without the unitary intensity introduced, the correlation among intensities may be very strong only with respect to many reflexion terms, while in the Patterson map the values at a few restricted points already have a strong correlation. This may reflect the fact that the Patterson method has been widely used while the direct method has been developed but gradually. In short, human brains full of insight, though not good at dealing with multi-variables, can fully utilize the correlation at local regions included, for instance, in the minimum function. On the contrary, a computer into which insight is not easy to introduce is a suitable tool for obtaining some useful information which can be deduced only from the simultaneous distribution of many variables.

### Comments on other studies from the present point of view

The following discussion uses the concepts of $\mathbf{I}^h$ space and $\mathbf{x}_j$ space. The $\mathbf{x}_j$ space is a Euclidean metric space with the periodicity of unity for each coordinate. In what follows, points which are identical with each other with modulus 1 are considered to be identical. It is often possible for more than one point in $\mathbf{x}_j$ space to correspond to an identical crystal. These points are usually transformed from each other by permuting coordinate axes corresponding to atoms of the same kind. Apart from these points, which will also be considered to be identical, points corresponding to the identical $\mathbf{I}^h$, if any, are homometric structures.

Any point in $\mathbf{x}_j$ space, even one corresponding to a fictional crystal, is always accompanied by a point in $\mathbf{I}^h$ space. On the other hand, a point in $\mathbf{I}^h$ space does not always correspond even to a fictional crystal. To take a trivial example, a set of $I^h$ values, some of which satisfy $I^h > 1$, apparently does not express any

Table 8. *Correlation between $V_{2/96}$ and $V_{3/96}$*

| $V_{2/96}$ \ $V_{3/96}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | | | | | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | 2 | 1 | | 1 | | |
| 10 | | | | | | | | | | | | | 1 | 3 | 38 | 38 | 5 | 6 | | | | |
| 12 | | | | | | | | | 1 | 1 | 34 | 59 | 148 | 104 | 79 | 30 | 1 | 1 | | | | |
| 14 | | | | | | 1 | 6 | 24 | 87 | 246 | 306 | 320 | 162 | 20 | 7 | 1 | | | | | | |
| 16 | | | | 3 | 6 | 45 | 211 | 475 | 573 | 447 | 279 | 69 | 10 | | | | | | | | | |
| 18 | | | 6 | 56 | 188 | 453 | 680 | 575 | 313 | 150 | 24 | 3 | | | | | | | | | | |
| 20 | 2 | 18 | 115 | 327 | 502 | 530 | 350 | 152 | 33 | 1 | | | | | | | | | | | | |
| 22 | 15 | 83 | 216 | 323 | 252 | 150 | 59 | 6 | 1 | | | | | | | | | | | | | |
| 24 | 36 | 103 | 140 | 92 | 20 | 7 | | | | | | | | | | | | | | | | |
| 26 | 29 | 38 | 39 | 7 | 1 | | | | | | | | | | | | | | | | | |
| 28 | 6 | 5 | 3 | 1 | | | | | | | | | | | | | | | | | | |
| 30 | | | | | | | | | | | | | | | | | | | | | | |
| 32 | | | | | | | | | | | | | | | | | | | | | | |

Table 9. *The information amount contained in* $V_{z/96}$ *(on the small set)*

| $z$ | $H(z)$ |
|---|---|
| 1 | 0·00 |
| 2 | 1·81 |
| 3 | 1·81 |
| 4 | 3·73 |
| 5 | 2·11 |
| 6 | 3·83 |
| 7 | 3·92 |
| 8 | 3·65 |
| 9 | 3·73 |
| 10 | 3·85 |
| 11 | 3·76 |
| 12 | 3·80 |
| 13 | 3·74 |
| 14 | 3·77 |
| 15 | 3·76 |
| 16 | 3·30 |
| 17 | 3·81 |
| 18 | 3·74 |
| 19 | 3·72 |
| 20 | 3·86 |
| 21 | 3·77 |
| 22 | 3·78 |
| 23 | 3·80 |
| 24 | 3·81 |
| 25 | 3·72 |
| 26 | 3·83 |
| 27 | 3·87 |
| 28 | 3·58 |
| 29 | 1·81 |
| 30 | 3·58 |
| 31 | 2·09 |
| 32 | 0·00 |
| 33 | 2·09 |
| 34 | 3·58 |
| 35 | 1·81 |
| 36 | 3·59 |
| 37 | 3·86 |
| 38 | 3·84 |
| 39 | 3·72 |
| 40 | 3·82 |
| 41 | 3·82 |
| 42 | 3·75 |
| 43 | 3·86 |
| 44 | 3·74 |
| 45 | 3·90 |
| 46 | 3·84 |
| 47 | 3·74 |
| 48 | 3·29 |

realizable crystal. The set of all meaningful points in $\mathbf{I}^h$ space, which will be denoted by $M$, cannot be outside a super-cube with length 1 for each edge in this space. If a point in $\mathbf{x}_j$ space is shifted along each of the $J$ coordinate axes by an infinitesimal distance, a corresponding point in $\mathbf{I}^h$ space shifts along each of $J$ directions also by an infinitesimal distance. Thus any movement in $\mathbf{x}_j$ space can be mapped in a sub-space of $J$ dimensions in $\mathbf{I}^h$ space, and the whole $\mathbf{x}_j$ space is mapped in $\mathbf{I}^h$ space as a sub-space $M$ of $J$ dimensions.

The extent to which the structure analysis has been well carried out is usually expressed by a parameter such as the $R$ index or $R$ factor, which corresponds to the distance between the observed point $I_{\text{obs}}$ and the calculated point $I_{\text{cal}}$ in $\mathbf{I}^h$ space. When the point $I_{\text{obs}}$ happens to belong to $M$, it is possible to make this distance zero. However, in general this would not happen.

The sub-space $M$ does not distribute uniformly in $\mathbf{I}^h$ space, being dense in some regions and disperse elsewhere. Two non-equivalent points more or less distant from each other in $\mathbf{x}_j$ space sometimes happen to have their mapping points very close to each other in $\mathbf{I}^h$ space; these structures are pseudohomometric, and it is difficult to tell which point represents the true structure, when the observed point falls near these two points. In such a case, more accurate observed intensities are required to determine the structure. On the other hand, when the relevant point falls in the sparsely populated region, even less accurate measurements will enable us to distinguish the true structure from a nearly similar but false one.

Suppose we plot in $\mathbf{x}_j$ space the value of the distance of $I_{\text{cal}}$ from the fixed point $I_{\text{obs}}$ in $\mathbf{I}^h$ space. If observed data are good enough, the true structure will agree with the point which has the minimum distance. This multi-dimensional map may be called 'an $R$ factor map' as used in the review articles (Hosoya, 1961, 1964). This idea seems to have been used in various investigations (for instance, Hosoya, 1958) and has extensively been studied by Milledge (1962). The $R$ factor map with contour lines drawn may have numerous peaks and troughs, and thus the $\mathbf{x}_j$ space may be divided into many regions, as it were, by a multi-dimensional network of watersheds. The number of these regions may be large but will certainly be far smaller than the number of points to be considered in the $\mathbf{x}_j$ space. Improvement of the Monte Carlo method of Vand & Niggli (1961) so as to make it practicable was effected only by introducing the optimal shift method (Niggli, Vand & Pepinsky, 1961), which is nothing but the Monte Carlo method for the above-mentioned regions.

Karle & Hauptman (1964) used $H'$ terms which are beyond the range of the observed $H$ terms, when they devised a method of refining the Patterson map. It may not be self-evident that it is allowable to use unobserved reflexion terms outside of the limiting sphere. However, this may be understood by the following considerations. When the projection onto a partial space with $H$ dimensions is given, the sub-space $M$ in $\mathbf{I}^h$ space with $H'$ dimensions more or less restricts the coordinates in certain regions, and, therefore, the values of $H'-H$ unobserved terms can be inferred to some extent.

In the above discussion it has been assumed that the distribution of the points in $\mathbf{x}_j$ space is uniform. But some regions in $\mathbf{x}_j$ space are entirely prohibited, for instance, because of steric hindrance due to the finite size of atoms. This fact may be clearly seen in a short note by Goedkoop, MacGillavry & Pepinsky (1951). Moreover, such a non-uniformity comes from coordination of atoms in covalent crystals, from Pauling's rules in ionic crystals or from the form of molecules in molecular crystals. In the example in the previous

chapter, $32^{96}$ points in $\mathbf{x}_j$ space were reduced to several tens of millions because of symmetry and steric hindrance. A set of the observed $I_{obs}$ values is not arbitrary but, as was pointed out in the above, more or less restricted. It is therefore favourable to take the weight of the points in $\mathbf{x}_j$ space into consideration. The results on SiC in the previous chapter showed various expectation values for reflexion intensities. This comes from the non-uniformity in $\mathbf{x}_j$ space, because both the big set and the small set include only those structures which satisfy certain limitations.

In relation to this discussion, it should be mentioned here that Hauptman (1964) showed how to infer the shape of the molecule directly from the intensity distribution without the information about phases. Conversely, when the shape of a molecule or a chemical unit of a crystal is known, a kind of unitary intensity, or the ratio of the observed intensities and the square of an absolute value of the structure factor for the above-mentioned unit, is very helpful to reduce the number of parameters $J$.

Let $M'$ denote the mapped points of $\mathbf{x}_j$ space limited by steric hindrance and other possible information; then $M'$ is a subset of $M$ and has more sparse distribution than $M$ has in $\mathbf{I}^h$ space, and therefore the requirement of experimental accuracy of $I_{obs}$ for determining the structure uniquely becomes less severe. The extreme case is seen in the structure analysis of 96*R*-SiC (Tokonami, 1966), in which only the qualitative observed values were used; on modification so as to make the structure point fall on $M'$, utilizing certain algebraic characters found in the Patterson function, the unique solution was obtained. In usual crystal analyses, it often happens that a few reflexions are considered to be subject to extinction effects and their intensity data are taken into account with small weights. This may be allowed from the viewpoint that only the points on $M'$ have meaning in $\mathbf{I}^h$ space.

The authors express their thanks to Prof. S. Miyake for his advice in improving the method of presentation of these concepts.

### References


DIAMOND, R. (1963). *Acta Cryst.* **16**, 627.

GOEDKOOP, J. A., MACGILLAVRY, C. H. & PEPINSKY, R. (1951). *Acta Cryst.* **4**, 491.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421.

HOSOYA, S. (1958). Dissertation, Univ. of Wales.

HOSOYA, S. (1961). *J. Cryst. Soc. Japan*, **3**, no. 1, 2 (in Japanese).

HOSOYA, S. (1964). *J. Cryst. Soc. Japan*, **6**, 56 (in Japanese).

JAGODZINSKI, H. (1949). *Acta Cryst.* **2**, 201.

KARLE, J. & HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 392.

KRISHNA, P. & VERMA, A. R. (1965). *Z. Kristallogr.* **121**, 36.

MILLEDGE, H. J. (1962). *Proc. Roy. Soc.* A, **267**, 566.

NIGGLI, A., VAND, V. & PEPINSKY, R. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p. 161. Oxford: Pergamon Press.

SAYRE, D. (1952). *Acta Cryst.* **5**, 843.

SHANNON, C. E. (1948). *Bell Syst. T. J.* **27**, 379.

TOKONAMI, M. (1966). *Miner. J. Japan*, **4**, 401.

TOKONAMI, M. & HOSOYA, S. (1965). *Acta Cryst.* **18**, 908.

VAND, V. & NIGGLI, A. (1961). In *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, p. 266. Oxford: Pergamon Press.


*Acta Cryst.* (1967). **23**, 25

# Prévision de quelques Images de Dislocations par Transmission des Rayons X (Cas de Laue symétrique)

PAR DANIEL TAUPIN

*Centre de Calcul Numérique, Laboratoire de Physique Théorique et Hautes Energies, Faculté des Sciences, 91-Orsay (Seine et Oise), France*

(*Reçu le* 12 *octobre* 1966)

The author's dynamical theory of the diffraction of *X* rays by distorted crystals is applied to the computation of the images obtained in the Laue case (transmission) when a single dislocation is contained in an otherwise perfect crystal. Some results which have been obtained for various orientations of the Burgers vector are presented: dislocation far from the upper surface (entry), the width of the incident beam being infinite; dislocation in the vicinity of the upper surface, the width of the incident beam being infinite; dislocation in the vicinity of the upper surface, the incident beam being very narrow. The results are in good agreement with topographs published by different authors.

Dans un précédent travail (Taupin, 1964*a*, *b*) nous avons montré comment, à partir des équations de Maxwell, on pouvait étendre la théorie dynamique, maintenant devenue classique, de la diffraction des rayons X par les cristaux parfaits au cas où les cristaux comportaient des défauts ou déformations élastiques, sans toutefois être limité à l'approximation en colonne comme dans la théorie de Howie & Whelan (1961,