

Science and Engineering Research Council

Rutherford Appleton Laboratory

CHILTON, DIDCOT, OXON, OX11 0QX

RAL-87-059

Ab Initio Crystal Structure Determination using an Image Reconstruction Technique

M W Johnson and W I F David

July 1987

Science and Engineering Research Council

The Science and Engineering Research Council does not accept any responsibility for loss or damage arising from the use of information contained in any of its reports or in any communications about its tests or investigations.

AB INITIO CRYSTAL STRUCTURE DETERMINATION
USING AN IMAGE RECONSTRUCTION TECHNIQUE

M W Johnson & W I F David

Rutherford Appleton Laboratory
Chilton
Didcot
Oxon OX11 0QX

July 1987

AB INITIO CRYSTAL STRUCTURE DETERMINATION

USING AN IMAGE RECONSTRUCTION TECHNIQUE

Abstract

A new algorithm, MEDIC (Maximum Entropy Direct Inversion of Crystallographic Data), is presented that provides a direct method of inverting the observed integrated intensities in a diffraction experiment to yield the scattering density within a unit cell. The method, based upon the METRIC algorithm [1], uses an image reconstruction technique employing the maximum entropy principle. It has been successfully tested on model systems with both centrosymmetric and non-centrosymmetric structures and on experimental neutron powder diffraction data from two small inorganic structures, $\text{Cu}_3(\text{PO}_4)_2$ and FeAsO_4 .

Introduction

The observation of the integrated intensities, I_{hkl} , of Bragg reflections and the derived magnitudes of the structure factors, $|F_{hkl}|$, does not lead directly to the solution of a crystal structure because the calculation of the scattering density $\rho(r)$ requires not only the magnitude but also the phase of the complex structure factor F_{hkl} .

Over the last 40 years sophisticated analytical methods have been developed that are based upon the probabilistic relationships between phases of different structure factors originally developed by Karle and Hauptman (1950) [2]. These methods have been considerably extended and have proved extremely successful in the determination of crystal structures. One of the more interesting developments over the past few years has been the application of maximum entropy (ME) techniques. These techniques, originally derived for underdetermined problems in information theory [3], and discussed extensively by Brice [4], show much promise in tackling the lack of knowledge associated with the absence of a-priori phase information. The recent work of

Gull, Livesey and Sira [5] exemplifies this approach that focuses on the need to determine the phases using combinations of the information provided by n-tuple phase relationships and the ME principle of maximising the objective function

$$S = - \sum \rho(r) \ln\{\rho(r)/\rho_0(r)\} \quad (1)$$

subject to the constraint ($\chi^2 \approx m$) imposed by the observed structure factors $|F_{hkl}^o|$:

$$\chi^2 = \sum_{k=1}^m \frac{(|F_{hkl}^o| - |F_{hkl}^c|)^2}{\sigma_{hkl}^2} \quad (2)$$

The summation in eq.1 is taken over the unit cell and $\rho_0(r)$ is an 'a priori' value for $\rho(r)$.

The maximum entropy approach proposed in the present paper differs from previous ME techniques in that it does not operate in Fourier space in the determination of phases but instead operates using the Patterson function in real space. The process thus proceeds in two stages:

$$\begin{array}{ccccc} \{|F_{hkl}| \} & \rightarrow & \{P(r)\} & \rightarrow & \{\rho(r)\} \\ i & & & & ii \end{array}$$

- (i) The Patterson map is generated either by standard Fourier transformation of $|F_{hkl}|^2$ or by ME methods [6].
- (ii) The Patterson map is inverted (decorrelated) to provide the real space structure.

The technique of Patterson decorrelation is discussed in the following section, and subsequent sections deal with the evaluation of the process using

both model and experimentally determined Patterson maps.

Patterson Map Decorrelation

The MEDIC algorithm consists of treating the Patterson map decorrelation as an image reconstruction problem in which the real space structure (the scattering density, $\rho(r)$, within a unit cell) represents the image to be reconstructed.

In the traditional image reconstruction problem the knowledge about the system consists of a set of observations C_1, C_2, \dots, C_n that represent a series of differently weighted linear sums over the image domain.

$$C_j = \sum_{i=1}^n w_{ij} \rho_i \quad (j = 1..m; m < n) \quad (3)$$

The problem is to obtain ρ_i given w_{ij} and C_j .

The MEDIC algorithm, too, casts the Patterson decorrelation as a linear problem. The Patterson map, $P(r)$, is subdivided into $N = n_x n_y n_z$ pixels, each of which are regarded as 'observations'. Thus, subdividing the scattering densities $\rho(r)$ on to a similar grid, one may write

$$P(r_j) = \sum_{r_i} \rho(r_i + r_j) \cdot \rho(r_i) \quad (4)$$

where the summation is performed over the unit cell.

Equation (2) is analogous to equation (1) if $\rho(r_i + r_j)$ and $\rho(r_i)$ are considered to be the weights and image respectively. If an iterative technique is adopted a self consistent solution for $\rho(r_i)$ may be found given that a suitable starting point is chosen and that $\rho(r_i + r_j)$ is adjusted at each iteration to agree with that of $\rho(r_i)$.

In the present work the starting point for $\rho(r_i)$ was taken to be a truncated version of the Patterson map such that

$$\begin{aligned}\rho(r_i) &= 1 \quad \text{if } (x_i < f, y_i < f, z_i < f \text{ and } \rho(r_i) > h) \\ &= 0.01 \quad \text{otherwise}\end{aligned}$$

where $r_i = x_i a + y_i b + z_i c$ (a, b, c are lattice vectors).

The choices of $f = 0.66$ and $h = 10$ ($P(r)$ is scaled so that $1000 < P(0) < 10000$) have been used throughout the present work and have not been adjusted to suit a particular problem.

The computer code MEDIC to perform this inversion is very concise and is freely available from the authors.

Evaluation

The MEDIC program was first tested on 'numerical Pattersons' created by numerical convolution of an original model density map, $\rho(r)$.

Density maps using 20^3 cells were first created from atomic positions by 'dressing' each position with a Gaussian density distribution and with a weight proportional to the atomic scattering factor. An example of the first four planes of FeAsO_4 so produced is shown in Figure 1.

The corresponding Patterson was formed by numerical integration

$$P(r_j) = \sum_{r_i} \rho(r_i + r_j) \rho(r_i) \quad (5)$$

and is shown in Figure 2.

The Patterson shown in Figure 2 was then inverted using the MEDIC program, a process which entailed no adjustable parameters or user intervention. The inverted result after 300K iterations is shown in Figure 3. Table 1 shows a list of the original atomic positions together with the reconstructed positions of the largest 24 peaks in the $\rho(r)$ map. These values were determined from the reconstructed density map by identifying as a peak any cell with 26 surrounding cells all possessing a lower scattering density. The peak coordinates and integrated intensity were obtained respectively by weighted average and summation over a 5-cell^3 volume. Since no phase

information has been included, the origin of the unit cell and the signs of their axes are not uniquely defined in the reconstructed density map. The reconstructed positions in Table 1 have been transformed to show the agreement obtained. A series of similar trials were undertaken on other compounds, the results of which are shown in Table 2. In all these cases the iterative procedure in MEDIC converged to the correct solution, which was then stable during further iterations. The final column in Table 2 shows the mean and maximum value of the distances between the initial and reconstructed atomic positions. For all the examples chosen, the maximum fractional distances between the model and reconstructed positions are much less than the pixel size of 0.05

The numerical reconstructions described above indicate the possibility of Patterson map decorrelation. However they do not address the fact that in any genuine ab initio structure determination the Patterson map will be inexact because the structure factors are both finite in number and subject to experimental error. Accordingly the method has been tested on two Patterson maps derived from experimentally recorded $\{|F_{hkl}|\}$ lists determined from neutron powder diffraction patterns of $\text{Cu}_3(\text{PO}_4)_2$ and FeAsO_4 . The use of a powder diffraction $\{|F_{hkl}|\}$ set means that structure factors of overlapping reflections were averaged - a process likely to hinder rather than aid the inversion.

$\text{Cu}_3(\text{PO}_4)_2$

Neutron diffraction data were recorded up to a maximum $\sin\theta/\lambda$ of 0.7\AA^{-1} on the HRPD powder diffractometer at the ISIS facility [7]. Although a total of 768 reflections lie within this range only 610 reflections were non-overlapping. In the present analysis the structure factors for reflections in overlapping groups were obtained by simply assigning the same, average value to each reflection within the group. To minimise Fourier truncation errors the Patterson map was averaged over a cube of 0.7\AA . All negative Patterson density was set to zero prior to the Patterson decorrelation.

The resulting Patterson map for $\text{Cu}_3(\text{PO}_4)_2$ is shown in Figure 4. The MEDIC program produced a separation of the true 13 atomic positions from noise in the ρ map after 16000 cycles. Further cycles of the program led to dramatically improved signal-to-noise such that after 60000 cycles the ratio

of the height of the lowest 'genuine' peak to the highest noise peak in the ρ map was 8000/1. This may be compared with the ratio of 'lowest signal/highest noise' of 2.4 obtained from the same set of structure factor magnitudes using the direct methods package, MITHRIL [8]. A list of the reconstructed xyz values for $\text{Cu}_3(\text{PO}_4)_2$ (obtained in the same manner as described above for the numerical Pattersons) is shown in Table 3 together with a set determined by least-squares refinement. It will be seen that the agreement is excellent and always less than the cell dimension of 0.05.

FeAsO_4

Neutron diffraction data were similarly recorded up to a maximum $\sin\theta/\lambda$ of 0.7\AA^{-1} on the HRPD powder diffractometer [7]. In this case a total of 901 reflections lie within the range although only 550 reflections were non-overlapping. Hence a total of 351 reflections lay in overlapping groups and were assigned average structure values.

The MEDIC program produced a separation of the true 24 atomic positions from the noise in the ρ map after 50,000 cycles. In all subsequent cycles up to 300,000 the ρ map produced a correct separation of 23 atomic positions (40% of maps) or 24 positions (60%). The 'lost' position in the maps which only separated 23 of the 24 positions varied from one map to another, and even using only these 'incomplete' maps the correct structure would be revealed by a simple correlation technique. A list of the reconstructed xyz values for FeAsO_4 (from one of the '24-position' maps) is shown in Table 4 together with a set determined by least-squares refinement. It will be seen that the agreement is again excellent.

Conclusion

An algorithm, based upon image reconstruction techniques, is presented that provides a new method of crystal structure determination. Although not yet demonstrated on a wide number of materials, the above examples were performed on powder data sets, where the problem is made harder by the lack of correct values for the structure factors of overlapping reflections. Even so it has

exhibited a remarkable stability and signal-to-noise capability.

Clearly the limits of the method need to be explored and, more importantly, the possibility of producing improved variants on the above method. The most important of these would be the use of the known crystal symmetry in constraining the density map $\rho(r)$. It should be emphasised that the present method imposes no constraints on $\rho(r)$ and the inclusion of symmetry constraints should greatly extend the range of the present method. Other possibilities include the use of ME Pattersons to overcome the problems of 'ripples' within $P(r)$, and the use of larger cell arrays.

References

1. Johnson M.W. (1987) Rutherford Appleton Laboratory Report **RAL-87-058**
2. Karle J. & Hauptman H. (1950) Acta Cryst. 3, 181-187.
3. Jaynes E.T. (1957) Phys. Rev. 106, 620-630.
4. Bricogne G. (1984) Acta Cryst. A40, 410-445.
5. Gull S.F, Livesey A.K. & Siva D.S. (1987) Acta Cryst. A43, 112-117.
6. David W.I.F. (1986) Rutherford Appleton Laboratory Report **RAL-86-066**
7. Johnson M.W. & David W.I.F. (1985) Rutherford Appleton Laboratory Report **RAL-85-112**
8. Gilmore C.J. (1984) J. Appl. Cryst. 17, 42-46.

MODEL			MEDIC			Δ
x	y	z	x	y	z	
0.173	0.462	0.763	0.173	0.462	0.763	0.000
0.327	0.962	0.737	0.328	0.964	0.736	0.002
0.827	0.538	0.237	0.828	0.541	0.236	0.002
0.673	0.038	0.263	0.673	0.043	0.263	0.004
0.073	0.202	0.223	0.071	0.203	0.229	0.006
0.427	0.702	0.277	0.426	0.705	0.279	0.003
0.927	0.798	0.777	0.926	0.799	0.782	0.005
0.573	0.298	0.723	0.573	0.300	0.726	0.004
0.257	0.101	0.420	0.255	0.100	0.420	0.002
0.243	0.601	0.080	0.242	0.604	0.080	0.003
0.743	0.899	0.580	0.743	0.898	0.580	0.000
0.757	0.399	0.920	0.756	0.401	0.922	0.002
0.027	0.377	0.384	0.027	0.380	0.384	0.002
0.473	0.877	0.116	0.475	0.875	0.112	0.004
0.973	0.623	0.616	0.973	0.623	0.616	0.000
0.527	0.123	0.884	0.528	0.121	0.882	0.003
0.889	0.077	0.152	0.889	0.083	0.153	0.006
0.611	0.577	0.348	0.612	0.577	0.347	0.001
0.111	0.923	0.848	0.111	0.925	0.849	0.002
0.389	0.423	0.652	0.392	0.421	0.650	0.004
0.125	0.267	0.939	0.122	0.264	0.943	0.005
0.375	0.767	0.561	0.374	0.770	0.560	0.003
0.875	0.733	0.061	0.872	0.733	0.062	0.002
0.625	0.233	0.439	0.625	0.238	0.440	0.005

Table 1

Structure	atoms/unit cell	Δ -mean	Δ -max
TiO ₂	6	0.0023	0.0059
TeO ₂	12	0.0054	0.0114
FeNbO ₄	12	0.0037	0.0051
Al ₂ O ₃	20	0.0080	0.0214
NaUO ₃	20	0.0057	0.0101
FeAsO ₄	24	0.0034	0.0062

Table 2

LEAST-SQUARES			MEDIC			Δ
x	y	z	x	y	z	
0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.278	0.226	0.316	0.286	0.205	0.339	0.032
0.722	0.774	0.684	0.717	0.781	0.697	0.015
0.359	0.353	0.779	0.358	0.346	0.787	0.011
0.641	0.647	0.221	0.643	0.655	0.239	0.019
0.846	0.344	0.339	0.846	0.349	0.350	0.012
0.154	0.656	0.661	0.149	0.672	0.671	0.019
0.332	0.651	0.169	0.332	0.655	0.183	0.014
0.668	0.349	0.831	0.672	0.346	0.832	0.005
0.230	0.227	0.005	0.213	0.229	0.014	0.019
0.770	0.773	0.995	0.762	0.770	0.010	0.017
0.379	0.150	0.633	0.369	0.161	0.636	0.014
0.621	0.850	0.367	0.619	0.860	0.371	0.010

Table 3

LEAST-SQUARES			MEDIC			Δ
x	y	z	x	y	z	
0.173	0.462	0.763	0.173	0.462	0.763	0.000
0.327	0.962	0.737	0.329	0.962	0.712	0.025
0.827	0.538	0.237	0.832	0.540	0.240	0.005
0.673	0.038	0.263	0.677	0.035	0.267	0.007
0.073	0.202	0.223	0.074	0.212	0.224	0.009
0.427	0.702	0.277	0.438	0.712	0.263	0.020
0.927	0.798	0.777	0.938	0.812	0.789	0.021
0.573	0.298	0.723	0.572	0.296	0.723	0.001
0.257	0.101	0.420	0.254	0.100	0.430	0.010
0.243	0.601	0.080	0.242	0.612	0.076	0.011
0.743	0.899	0.580	0.738	0.912	0.589	0.016
0.757	0.399	0.920	0.755	0.412	0.921	0.012
0.027	0.377	0.384	0.013	0.373	0.375	0.017
0.473	0.877	0.116	0.488	0.889	0.126	0.021
0.973	0.623	0.616	0.981	0.612	0.621	0.014
0.527	0.123	0.884	0.521	0.115	0.885	0.009
0.889	0.077	0.152	0.890	0.074	0.174	0.022
0.611	0.577	0.348	0.593	0.564	0.321	0.035
0.111	0.923	0.848	0.102	0.912	0.846	0.014
0.389	0.423	0.652	0.388	0.412	0.621	0.032
0.125	0.267	0.939	0.138	0.262	0.936	0.014
0.375	0.767	0.561	0.358	0.762	0.569	0.019
0.875	0.733	0.061	0.855	0.744	0.067	0.023
0.625	0.233	0.439	0.638	0.222	0.443	0.017

Table 4.


```

213114 13 1 1 1 0 0 0 0 0 0 0 0 0 1 1 1 13114
106 57 7 1 1 1 1 2 2 1 0 1 2 2 1 1 1 1 7 57
13 7 1 3 3 1 3 12 16 7 1 7 16 12 3 1 3 3 1 7
0 0 0 2 3 1 5 23 35 18 4 18 35 23 5 1 3 2 0 0
0 0 0 1 2 2 3 13 24 15 4 15 24 13 3 2 2 1 0 0
0 0 0 2 9 10 4 4 7 5 2 5 7 4 4 10 9 2 0 0
2 1 0 3 16 22 10 2 1 1 0 1 1 2 10 22 16 3 0 1
3 2 0 2 16 24 12 2 1 0 0 0 1 2 12 24 16 2 0 2
2 2 2 3 6 7 4 2 1 0 0 0 1 2 4 7 6 3 2 2
3 4 10 16 11 2 4 7 4 1 0 1 4 7 4 2 11 16 10 4
3 6 19 33 21 4 7 14 8 1 0 1 8 14 7 4 21 33 19 6
3 4 10 16 11 2 4 7 4 1 0 1 4 7 4 2 11 16 10 4
2 2 2 3 6 7 4 2 1 0 0 0 1 2 4 7 6 3 2 2
3 2 0 2 16 24 12 2 1 0 0 0 1 2 12 24 16 2 0 2
2 1 0 3 16 22 10 2 1 1 0 1 1 2 10 22 16 3 0 1
0 0 0 2 9 10 4 4 7 5 2 5 7 4 4 10 9 2 0 0
0 0 0 1 2 2 3 13 24 15 4 15 24 13 3 2 2 1 0 0
0 0 0 2 3 1 5 23 35 18 4 18 35 23 5 1 3 2 0 0
13 7 1 3 3 1 3 12 16 7 1 7 16 12 3 1 3 3 1 7
106 57 7 1 1 1 1 2 2 1 0 1 2 2 1 1 1 1 7 57

108 58 7 1 0 0 0 0 0 0 0 1 1 0 1 8 10 4 7 58
54 29 4 2 3 1 1 2 2 0 0 2 3 2 1 6 7 3 4 29
7 4 2 9 11 4 3 8 8 3 1 9 18 11 2 2 3 1 1 4
1 0 2 9 12 5 4 13 18 9 4 21 39 22 3 0 0 0 0 1
0 0 0 3 6 4 2 8 16 11 4 14 23 13 2 0 0 0 0 1
1 1 0 2 9 10 4 4 7 6 2 4 6 3 2 4 3 0 0 0
5 4 2 3 12 15 7 5 3 1 1 4 6 4 7 13 9 1 0 2
7 6 2 2 7 11 7 6 3 1 2 5 6 4 9 18 12 2 1 3
6 3 1 1 2 3 3 3 2 1 2 4 4 4 4 6 6 4 5 7
6 1 2 4 3 1 4 8 5 1 1 2 8 14 7 4 11 17 18 16
8 1 4 7 5 2 8 16 10 1 0 2 12 23 11 6 20 32 29 21
6 1 2 4 3 1 4 8 5 1 1 2 8 14 7 4 11 17 18 16
6 3 1 1 2 3 3 3 2 1 2 4 4 4 4 6 6 4 5 7
7 6 2 2 7 11 7 6 3 1 2 5 6 4 9 18 12 2 1 3
5 4 2 3 12 15 7 5 3 1 1 4 6 4 7 13 9 1 0 2
1 1 0 2 9 10 4 4 7 6 2 4 6 3 2 4 3 0 0 0
0 0 0 3 6 4 2 8 16 11 4 14 23 13 2 0 0 0 0 1
1 0 2 9 12 5 4 13 18 9 4 21 39 22 3 0 0 0 0 1
7 4 2 9 11 4 3 8 8 3 1 9 18 11 2 2 3 1 1 4
54 29 4 2 3 1 1 2 2 0 0 2 3 2 1 6 7 3 4 29

13 7 1 0 0 0 0 0 0 0 1 5 6 2 4 24 28 10 2 7
7 4 1 2 3 1 0 0 0 0 2 8 10 4 4 18 22 8 1 4
2 1 2 11 15 6 1 2 2 1 2 9 13 6 2 7 8 3 1 2
4 1 2 14 21 10 2 3 4 3 3 11 17 9 1 1 1 1 4 7
6 1 1 7 11 6 1 2 5 5 4 6 10 5 1 0 0 1 5 9
3 1 1 3 5 4 2 3 4 3 2 3 4 2 1 1 1 1 2 4
8 8 7 6 4 5 8 13 8 2 3 9 13 7 2 3 3 1 1 3
13 12 9 7 3 3 10 19 12 3 8 16 15 7 2 4 6 5 2 5
9 5 3 3 2 2 5 10 8 4 8 12 8 5 3 5 10 11 9 10
9 1 1 4 6 3 3 6 6 4 3 4 7 12 10 9 10 11 19 23
11 1 1 5 8 5 6 10 8 3 1 2 11 20 16 13 12 13 25 29
9 1 1 4 6 3 3 6 6 4 3 4 7 12 10 9 10 11 19 23
9 5 3 3 2 2 5 10 8 4 8 12 8 5 3 5 10 11 9 10
13 12 9 7 3 3 10 19 12 3 8 16 15 7 2 4 6 5 2 5
8 8 7 6 4 5 7 13 8 2 3 9 13 7 2 3 3 1 1 3
3 1 1 3 5 4 2 3 4 3 2 3 4 2 1 1 1 1 2 4
6 1 1 7 11 6 1 2 5 5 4 6 10 5 1 0 0 1 5 9
4 1 2 14 21 10 2 3 4 3 3 11 17 9 1 1 1 1 4 7
2 1 2 11 15 6 1 2 2 1 2 9 13 6 2 7 8 3 1 2
7 4 1 2 3 1 0 0 0 0 2 8 10 4 4 18 22 8 1 4

0 0 0 1 1 0 0 1 1 1 6 17 17 5 3 15 18 6 1 0
0 0 1 4 5 2 0 1 0 1 6 22 24 8 3 12 14 5 1 0
3 0 2 13 18 8 1 0 0 0 3 12 15 5 1 5 5 2 2 4
13 2 2 16 26 13 1 0 1 2 3 4 5 2 1 3 3 2 11 21
17 2 1 9 16 8 1 0 1 5 7 3 1 1 2 4 3 2 15 29
7 2 3 4 3 2 1 2 2 2 3 2 2 1 1 4 6 3 6 11
4 5 11 11 3 1 7 14 9 2 3 7 9 4 1 6 11 6 2 2
5 7 14 13 3 2 13 25 17 5 9 14 11 5 1 5 12 9 2 2
3 3 4 6 4 3 10 19 19 13 11 12 6 3 3 6 14 14 6 3
3 1 1 9 15 8 4 9 18 17 7 3 3 6 15 15 10 7 7 7
3 0 1 12 20 11 2 4 11 10 3 1 3 11 25 24 9 4 7 9
3 1 1 9 15 8 4 9 18 17 7 3 3 6 15 15 10 7 7 7
3 3 4 6 4 3 10 19 19 13 11 12 6 3 3 6 14 14 6 3
5 7 14 13 3 2 13 25 17 5 9 14 11 5 1 5 12 9 2 2
4 5 11 11 3 1 7 14 9 2 3 7 9 4 1 6 11 6 2 2
7 2 3 4 3 2 1 2 2 2 3 2 2 1 1 4 6 3 6 11
17 2 1 9 16 8 1 0 1 5 7 3 1 1 2 4 3 2 15 29
13 2 2 16 26 13 1 0 1 2 3 4 5 2 1 3 3 2 11 21
3 0 2 13 18 8 1 0 0 0 3 12 15 5 1 5 5 2 2 4
0 0 1 4 5 2 0 1 0 1 6 22 24 8 3 12 14 5 1 0

```

Z = 0.0

Z = 0.05

Z = 0.10

Z = 0.15

Fig. 2

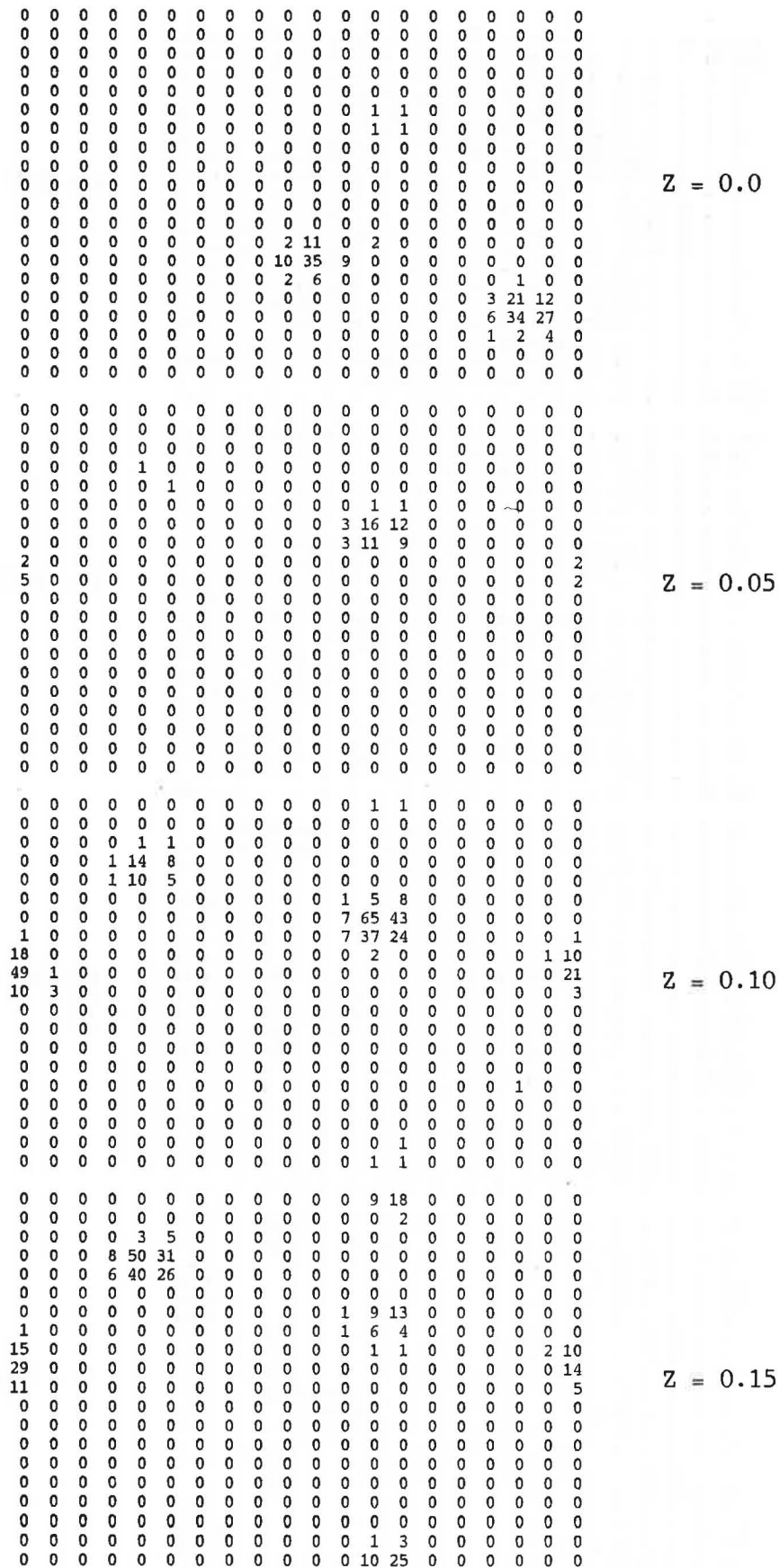


Fig. 3

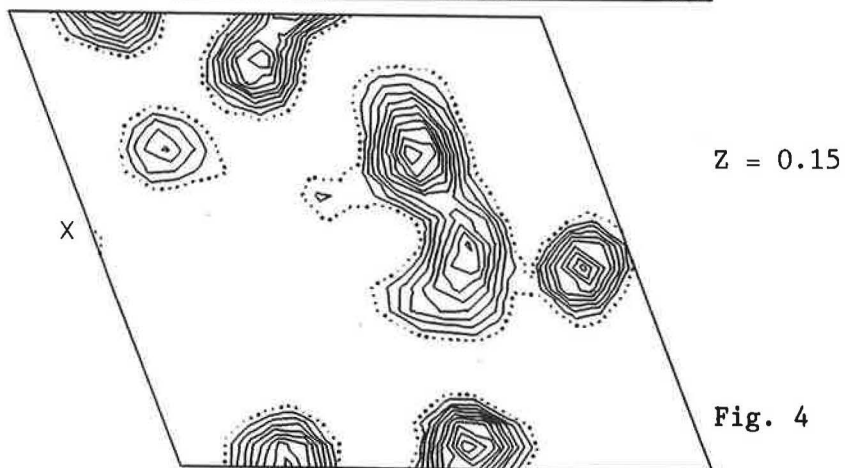
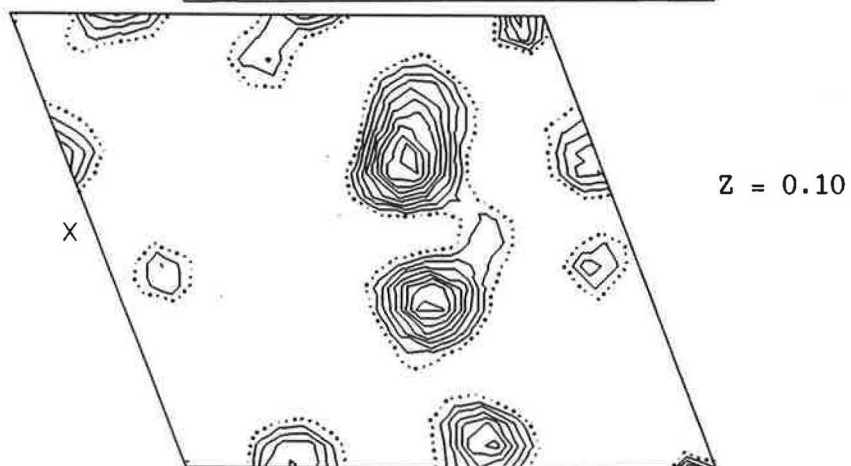
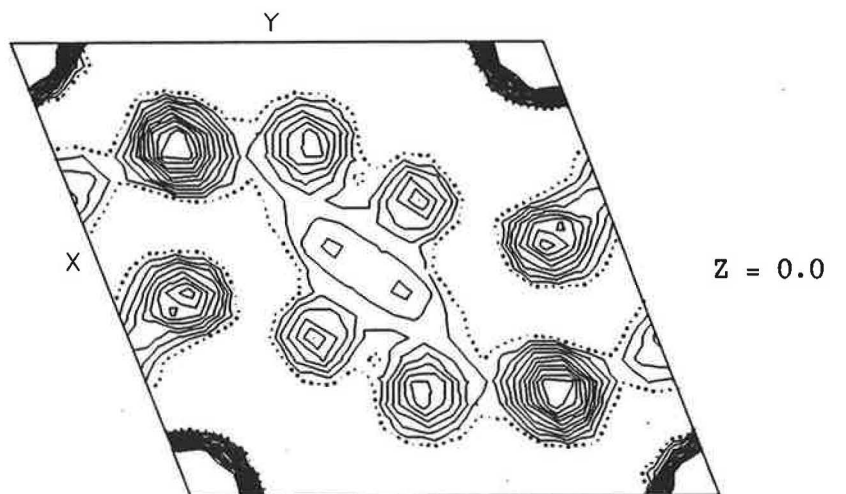


Fig. 4

1A

