

The Direct Determination of Centrosymmetric Electron Density Distributions by Maximizing the Integrated Cube of the Electron Density

BY ERIC STANLEY

Department of Physics, University of New Brunswick, Saint John, New Brunswick, Canada

(Received 30 October 1978; accepted 16 July 1979)

Abstract

The two characteristics of all fully refined electron density distributions are that they do not go negative and the electron density is concentrated into a small number of peaks. The correct set of structure factor signs is defined as a set which gives rise to an electron density distribution which has these characteristics. A simple function, sensitive to these characteristics, is the integrated cube of the electron density; it can distinguish between positive and negative peaks and between a 'peaky' and a dispersed distribution. An electron density map is calculated using only the known signs. Then each structure factor's contribution to the map is added with both alternative signs, in order of magnitude. The correct sign is that giving rise to the larger value of the integrated cube of the electron density. The term is added with the correct sign and the next structure factor similarly dealt with. A recycling procedure is also described. The method of refinement has been tried on several known two- and three-dimensional structures with great success enabling the electron density to be fully refined without recourse to any model. The same criterion can be used to determine the signs of the initial set of structure factors. Electron density distributions are calculated for all permutations of signs of the initial set. The particular combination chosen as correct is that combination giving the largest value of the integrated cube of the electron density. In all the cases tried the method gave the correct signs.

Introduction

Although direct methods have been used for many years to obtain the initial phases (or signs) of the stronger structure factors, the refinement of a structure has been accomplished with a model. Positional and thermal parameters are adjusted systematically using the agreement between observed and calculated structure factors as the criterion of refinement. Least squares, Fourier difference and minimum residual methods are examples of this approach. Gassman (1975, 1976) has proposed methods of refinement in

real space rather than in reciprocal space, Podjarny & Yonath (1977) have suggested a method for extended phase determination and Simonov (1969) has discussed the determination of phases in real space. It is proposed, in this paper, to refine the electron density distribution itself without the aid of any model or preconception of the nature of the structure.

The correct electron density distribution

An examination of the electron density distributions of fully refined structures reveals two main characteristics. The electron density is mainly confined to a 'small' number of isolated peaks on a background that seldom goes very negative. There are obvious reasons for such a distribution if it corresponds to a real crystal structure.

We can define the correct set of signs of the structure factors as a set which gives rise to an electron density distribution having just these characteristics. In looking for a suitable figure of merit with which to test an electron density distribution for both 'peakiness' and 'non-negativity', we must consider functions which can discriminate between peaky and dispersed distributions and can also distinguish between positive and negative peaks. Any integral power of the electron density greater than one can distinguish between peaky and dispersed distributions and any odd power of the electron density can distinguish between positive and negative peaks. A simple function which can distinguish both of these characteristics is the third power. For this reason the function maximized is the integrated cube of the electron density. Any function which is antisymmetric, or even asymmetric about zero could be used. As in all methods of sign determination, there is no guarantee that the solution is unique. The results indicate that the method does result in the direct evaluation of the electron density distribution. There is, of course, the ever-present possibility of refining to a homometric distribution but this has not yet been observed. In the final analysis it is the chemical plausibility that determines whether or not a structure has been correctly determined.

Application

The signs that can be allocated to the larger structure factors by the position of a heavy atom or atoms, by direct methods *etc.* are assigned. It is desirable to include terms from different seminvariant groups in order to fix the origin. Using just these known signs, an electron density map is calculated. Although Gassman (1977) has studied some of the information-theory aspects of structure determination, even quite elementary considerations lead to the conclusion that the largest structure factors contain the most information. For this reason, the remainder of the structure factors are tested in order of magnitude, one at a time, by adding their contribution to the electron density with alternative signs. The value of $\sum \rho^3$ is calculated with each sign. The sign giving the larger value of $\sum \rho^3$ is assumed to be correct and the term is added to the electron density map. The next structure factor is dealt with in the same way. Ideally, when this has been completed for all terms, all the signs should be correct.

Recycling data in refinement procedures is a familiar process in crystal structure analysis. A recycling procedure can be adopted in this method by going through the structure factors again and observing the effect on $\sum \rho^3$ as the sign of each term is changed. If the same order of the terms is used there seldom seems to be many changes. However, if the order of the structure factors is changed from the original order of magnitude to the order of magnitude of the difference in $\sum \rho^3$ when

the sign is changed, refinement of the electron density distribution can continue (*i.e.* in order of sensitivity of $\sum \rho^3$ to the sign change). Each time a sign is indicated to be 'incorrect', the term is replaced by the term having the 'correct' sign in the electron density map.

In the initial testing stages of the method, structure factors and normalized (or unitary) structure factors were both tried. It was found that the results were markedly better when the structure factors were used. Presumably this is because of the definition of the correct distribution and the criterion used to test the distributions. *E* or *U* maps, as is commonly observed, have many false peaks and deep negative areas because, although the *E*'s correspond to point atoms, the distribution is calculated using a very limited series: *i.e.* series termination errors are much more acute with *E*'s than with *F*'s.

Testing the method

The method was tested on a random sample of known structures. Those used were those for which the author had the data available either because he was involved in the original determination or because he happened to have a reprint or other access to the data.

1. An artificial structure

The artificial two-dimensional structure sketched in Fig. 1(a) was used for the initial test. Structure factors

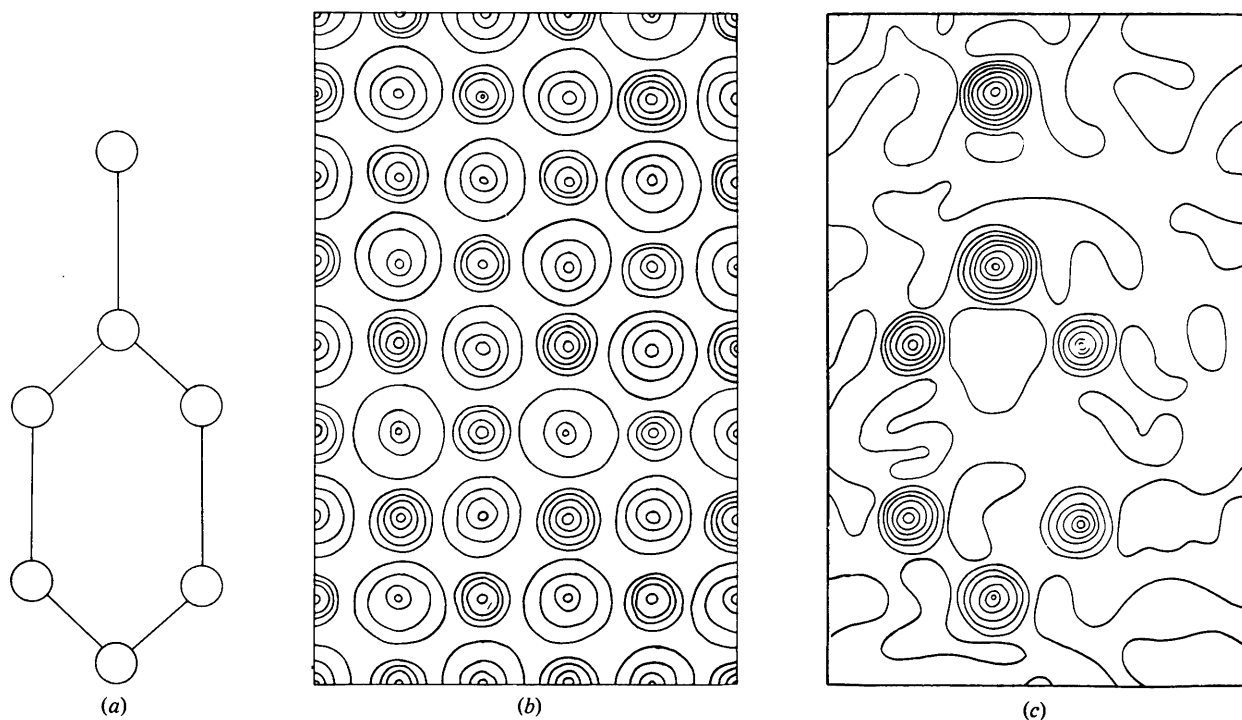


Fig. 1. Artificial structure. (a) The model, (b) the electron-density distribution calculated from seven known signs, (c) the electron density calculated after refinement.

were calculated for the structure and the signs of the first few strong structure factors were assigned. Initially seven signs only were assigned. In the initial determination 142 out of 161 signs were correct and after one recycle all the signs were correct. Figs. 1(b) and (c) show the initial and final electron density maps, respectively. Because of the success it was decided to reduce the number of known signs. With as few as three known signs the structure will refine to all correct signs after three recycling stages. Inadvertently, the three used were insufficient to define uniquely the origin. However, the method is capable of choosing its own origin presumably because of the accidental order in which the successive terms are introduced.

2. Tin monochloroacetate

The structure of this material has been determined by Faggiani, Johnson, Brown & Birchall (1978). With the nine strongest structure factors, the refinement of the electron density was undertaken for the [010] projection. The electron density map after the initial sign determination did not change on refinement. The number of terms assigned signs was reduced successively. With only three known signs the electron density map was fully refined in the first cycle and would not refine further. The authors did not give the signs of their calculated structure factors but the electron density map appears to be correct. The progress of the refinement is given in Fig. 2.

3. Triphenyl phosphate

The structure of triphenyl phosphate has been determined by Davies & Stanley (1962). The signs of twenty strong reflexions in the [010] projection were assigned

from the position of the phosphorous atom. In the initial sign determination, 200 out of 300 signs were correct and after recycling the number of wrong signs was reduced to 40. The progress of refinement is shown in Fig. 3 and compares well with the published refined electron density. The number of assigned signs was reduced successively in order to test the method more thoroughly. As few as five correct initial signs were necessary to enable refinement to take place.

4. Other structures

In addition, the method was used on the [010] projection of diphenylene naphthacene (Bennett & Hanson, 1953), on the [001] projection of Roussin's red ethyl ester (Thomas, Robertson & Cox, 1958) used by Woolfson (1961) to demonstrate sign relationships, and also on the three-dimensional data on the same material.

In all cases, the signs of the larger terms were all correctly determined; it is only the weakest of the structure factors whose sign may be incorrect. This is not surprising since the effect of the weaker terms on the electron density distribution is quite small.

The time involved for the calculation is of interest. With programs written in Fortran on the IBM 3032, the two-dimensional refinements took about 10 s per cycle per 100 terms.

A summary of the results is given in Table 1.

How many signs are required to define a structure?

It is clear that a structure is determined by a very few strong terms even if it is unrecognizable. If a heavy atom is present the number of required terms is very

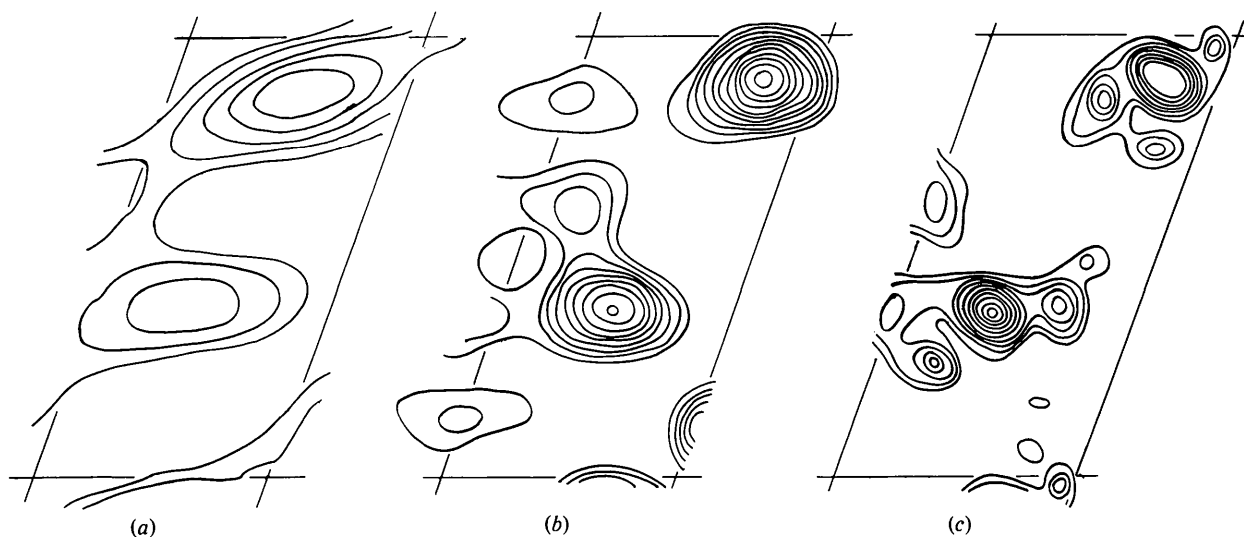


Fig. 2. Tin monochloroacetate. (a) The electron density calculated from three known signs, (b) the electron density calculated from nine known signs, (c) the refined electron density.

Table 1. Summary of the tests of refinement procedures on six structures

Structure	Dimensions	No. of positional parameters	Total no. of terms	No. of signs originally assigned	No. correctly determined in 1st cycle	No. correct after refinement
Artificial structure	2	14	161	3	142	161
Tin monochloroacetate	2	20	108	3	108	—
Triphenyl phosphate	2	46	300	5	200	260
Diphenylene naphthacene	2	30	77	20	65	67
Roussin's red ethyl ester	2	16	81	11	63	69
Roussin's red ethyl ester	3	24	125	14	101	118

Table 2. Roussin's red ethyl ester signs of 81 structure factors determined in order of magnitude of $F(hkl)$

Assigned signs	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Signs correct in initial cycle	40	41	48	49	50	50	55	60	63	64	68	67	69	70
% correct	50	52	61	62	63	63	69	75	80	81	86	85	87	89
First incorrect sign in order	2	4	6	6	6	12	12	15	19	19	43	43	43	58
in magnitude of $ F(hkl) $	105	65	63	63	65	45	45	38	34	34	16	16	16	8

small indeed. Even the 23 atom structure of triphenyl phosphate needed only five known signs to enable an interpretable electron density map to be prepared by this method of refinement. In the case of tin monochloroacetate, which contains two heavy atoms, only three strong reflexions were required. In the absence of a heavy atom, as in diphenyl naphthacene, and lacking any large unitary structure factors, many more terms are required for the initial electron density map. Even so the number of terms required is considerably less than the number of parameters.

A more detailed examination was conducted on one of the structures, Roussin's red ethyl ester. The results are shown in Table 2. There is a sharp discontinuity in the number of correctly determined signs and the magnitude of the first incorrect sign which, in this case, is at eleven known signs. When only very few known signs are used there is a danger of convergence on the false solution with a single large peak at one of the possible origins. This can be avoided by omitting the

regions in the neighbourhood of the origins from the integration.

Determination of the initial set of signs

It is worth considering whether the method of maximizing $\sum \rho^3$ will yield a correct initial set of signs, by calculating the electron density distributions for all permutations of the signs of the initial set of structure factors and selecting the set giving the greatest value of $\sum \rho^3$. The most probable set based on this criterion is, of course, the Patterson-like solution with all the signs positive (or other systematic combination of signs converging on one of the alternative origins). This catastrophe can be averted by omitting from the calculation the regions near the origins. The calculation of a large number of electron density distributions is a formidable task. The calculation can be reduced considerably if it is arranged that each successive dis-

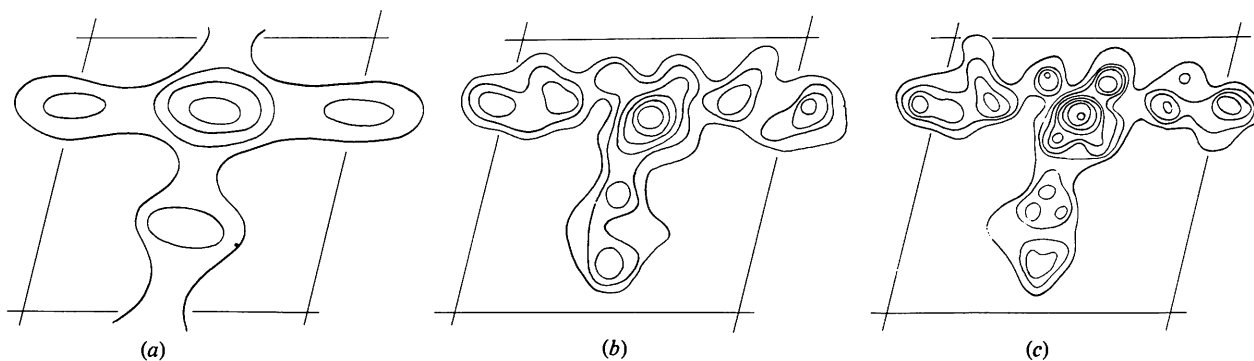


Fig. 3. Triphenyl phosphate. (a) The electron density calculated from five known signs, (b) the electron density calculated from 20 known signs, (c) the refined electron density.

tribution differs from the previous distributions by the change of sign of only one term. This can be achieved by adopting the Gray Code sequence, which is a sequence of binary numbers in which each number differs from the previous number by only one bit.

This method of assigning signs to an initial set of structure factors was tried using $F(000)$, two arbitrarily assigned signs from different seminvariant groups to define the origin, together with the fifteen strongest structure factors whose signs were permuted. In all the two-dimensional distributions refined, the artificial structure, triphenyl phosphate, diphenylene naphthacene, tin monochloroacetate and Roussin's red ethyl ester, the correct set of signs was produced.

The computer time on an IBM 3032 using Fortran was three and a half minutes.

Relationship to other direct methods

The electron density cubed can be written as

$$[\rho(x,y,z)]^3 = \left[\sum_h \sum_k \sum_l F(hkl) \cos 2\pi(hx + ky + lz) \right]^3.$$

In order that the integrated value of ρ^3 is large, it is necessary that many of the terms must be large and positive. One solution is the Patterson-like solution with all the F 's positive, or other systematic combination of signs. Other solutions will require some of the triple products of $F_1 F_2 F_3$ (where F_1, F_2 and F_3 are all large) to be positive. If all the triple products are positive, of course, we get the Patterson-like solution. One set of triple products is any closed set $F(\mathbf{h}), F(\mathbf{h}^1), F(\mathbf{h} \pm \mathbf{h}^1)$. Thus, amongst the triple products in the ρ^3 calculations are included all the sign relationships. There are many other triple products which are not used in the sign relationship expression.

It was noted by Sayre (1953) that

$$\int \rho^3 dv = \sum_h \sum_k F_h F_k F_{-h-k},$$

but it was not observed that the correct set of signs corresponding to a real structure would maximize this function.

Conclusion

It has been demonstrated that direct refinement of the electron density by the maximizing of $\sum \rho^3$ in both two and three dimensions is possible without ever proposing a model or calculating a structure factor. It is suggested that the main uses of the method may be in cases where the known signs do not yield a recognizable electron density map, or to determine which of

several possible initial sets of signs is the most likely to be correct. Another possible use might be in identifying non-carbon atoms in organic structures from the peak heights alone without distorting the refinement process by assigning non-carbon atoms in a model during refinement.

Although $\sum \rho^3$ has been used as the discriminating function, it is possible that other functions including higher odd powers, even powers together with the sign, exponential and sinh functions, or some other function, perhaps related to peak shape, would be better.

In principle, at least, it should be possible to extend the method to non-centrosymmetric electron density distributions but the computing time involved might be prohibitive. Two possible approaches seem to be feasible. In the first approach, each term could be added to the initial electron density map with several different phases (say every 60°) and the $\sum \rho^3$ criterion used for selecting the best phase. Another possible approach would be to calculate only the real part of the map using a variety of fractions, A , of the structure factor and with either sign. This map will contain the structure and its centrosymmetrically related structure but should conform to the criteria used in this method. The imaginary parts of the structure factors will now have been determined in magnitude, since $A^2 + B^2 = F^2$ but needs to be assigned a sign, which can be done by the methods described in this paper.

I am indebted to the National Research Council for the award of grants during the tenure of which this work was done, and to the University of New Brunswick for the use of facilities of the Saint John Campus Computer Centre, to John Perrin who helped with the programming and the computing and to Dr Larry Garey and Dr David Carson for many productive discussions.

References

- BENNETT, A. & HANSON, A. W. (1953). *Acta Cryst.* **6**, 736–739.
 DAVIES, W. O. & STANLEY, E. (1962). *Acta Cryst.* **15**, 1092–1097.
 FAGGIANI, R., JOHNSON, J. P., BROWN, I. D. & BIRCHALL, T. (1978). *Acta Cryst.* **B34**, 3742–3743.
 GASSMAN, J. (1975). *Acta Cryst.* **A31**, 825–831.
 GASSMAN, J. (1976). *Acta Cryst.* **A32**, 274–280.
 GASSMAN, J. (1977). *Acta Cryst.* **A33**, 474–479.
 PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* **A33**, 655–661.
 SAYRE, D. (1953). *Acta Cryst.* **6**, 430–431.
 SIMONOV, V. I. (1969). *Acta Cryst.* **B25**, 1–4.
 THOMAS, J. T., ROBERTSON, J. H. & COX, E. G. (1958). *Acta Cryst.* **11**, 599–604.
 WOOLFSON, M. M. (1961). *Direct Methods in Crystallography*. Oxford Univ. Press.