

Note

A Procedure for the Construction of Voronoi Polyhedra

1. INTRODUCTION

There has been recently an increased interest in the use of Voronoi polyhedra as a geometrical tool for investigating the structures and properties of a variety of condensed non-crystalline systems. In addition to the earliest work which attempted on the basis of the equivalent Voronoi array, to develop characterisations of the disorder in liquids that could be translated into statistical mechanical terms [1-8], a number of other applications to a variety of systems have been reported. Kiang [9] used the method in discussing the distribution of interstellar matter, while Rahman [2] and Finney [6] calculated the Voronoi statistics of the ideal gas in order to check a theoretical model which was also applicable to grain growth in metals [10]. Polyhedral statistics have been used very extensively in structural studies of metallic glasses and the glass transition [11-16], with recent stress on characterising the possible onset of crystallisation in the soft core model [17] (J. N. Cape, L. V. Woodcock, and J. L. Finney, in preparation), and examining the structural changes across the crystal-melt interface [18, 19]. Recent applications to large, inhomogeneous systems include the quantification of packing, environment, molecular area, and solvent accessibility of proteins [20-25] and nucleic acids (G. Zaccai, B. Jacrot, A. V. Westerman and J. L. Finney, in preparation). Work has also continued using polyhedral statistics as characterisations of both liquids and crystal structures [26-28].

In view of the extension of the technique, efficient methods of computing the equivalent Voronoi polyhedron array have become of prime importance. The only algorithm to be published in detail [29] attempts comparisons with other methods. We describe here the detailed operation of the methods we developed in 1966. This algorithm, which has been applied successfully to homogeneous and heterogeneous assemblies with up to 8,000 points, has not been fully described previously in the literature.

2. THE ALGORITHM

The Voronoi polyhedron associated with a given "centre" i , in an assembly of N "centres" is defined as that volume of space containing all points closer to i than to any other center j . Thus the bounding surfaces are planes drawn perpendicular to intercentre vectors ij at their midpoints; the intersections of these planes form the polyhedron edges and vertices. In contrast to the approaches of Brostow *et al.* [29] and Richards [20], who first define the polyhedron faces, we obtain as a first step the

polyhedron vertices. As every point on each bounding plane associated with a given pair of centres ij is by definition equidistant from both i and j , the intersection of (in general) three planes ij, jk, kl is a point equidistant from the four defining points ($ijkl$). This is a polyhedron vertex whose cartesian coordinates (xyz) are the solutions of the set of simultaneous quadratic equations

$$(x_p - x)^2 + (y_p - y)^2 + (z_p - z)^2 = r^2 \quad (p = i, j, k, l) \dots \quad (1)$$

r is the (identical) distance between each centre (i, j, k, l) and the associated vertex. This set of equations reduces to three linear equations in (xyz) which are easily solved. A valid vertex must be closer to the centre ($ijkl$) than to any other centre in the assembly; the distances between the solution to (1) and a subset of the total assembly which could possibly contain any closer centres are calculated and checked against r before a vertex is accepted (see below).

In moving from the vertices to the polyhedron faces and edges, use is made of the index numbers of the four centres ($ijkl$) associated with each vertex. All ($ijkl$) combinations for a given centre i are listed, there being one set to each vertex. Centre i is common to all vertices associated with polyhedron i and is now ignored. We are thus left with M sets of triplets ($j_q k_q l_q$) ($q = 1, M$), one for each vertex r_q . All vertices common to a given face will have an index in common, which will refer to the neighbouring centre which together with i defines this face. Removing this index leaves M index pairs associated with a single face. These index pairs are then ordered cyclically starting from any vertex; two neighbouring vertices joined by an edge will have two indices in common.

A simple example of this vertex sorting process is illustrated by the 8-faced polyhedron in Fig. 1, for which the vertex indices are listed in Table 1. Centre 1 is contained within this polyhedron; each face is labelled with the index number of the neighbouring centre which shares the face, while each vertex is labelled with the index numbers of the four centres from which it is equidistant. Thus the label 1 occurs in every vertex in Table 1. Vertices 1 to 5 have index 2 in common and thus are associated also with the face labelled 2. The remaining index pairs have each index in common with

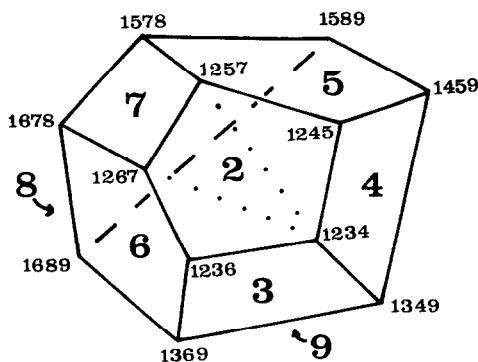


FIGURE 1

TABLE I
Vertex Index Table for the Polyhedron of Fig. 1

Vertex number <i>M</i>	Four centres associated with the vertex <i>M</i>			
1	1	2	3	4
2	1	2	3	6
3	1	2	4	5
4	1	2	5	7
5	1	2	6	7
6	1	3	4	9
7	1	3	6	9
8	1	4	5	9
9	1	5	7	8
10	1	5	8	9
11	1	6	7	8
12	1	6	8	9

another vertex of the same face. Thus starting with vertex 1 we can immediately order the vertices 12543 along neighbouring edges by passing to a vertex with a common index. Each index remaining in these pairs must occur at least twice, or the polyhedron face will not close (see below). An index occurring more than twice indicates a degenerate vertex equidistant from more than the minimum four centres. Such cases rarely occur by chance, but are found frequently in crystalline or partly-crystalline assemblies, and are dealt with as follows.

When two vertices are closer than a preset minimum distance R_{\min} (which may be set with respect to, e.g., computer rounding errors or expected limited accuracy of the data) they are replaced by a single "composite vertex" labelled with the set of five (rarely more) nonidentical indices which is the union of the index labels of the original vertices. The procedure continues as before, though only two of the three (rarely more) vertices remaining after removal of the two common face indices are used in tracing out the edge ordering. The remaining vertex (rarely vertices) is ignored. This procedure is particularly useful if the assembly is close to a simpler assembly —e.g., crystalline with small deformations. By setting R_{\min} artificially large, the crystallinity, which would otherwise be partly masked by the polyhedron complexity introduced by the deformations, becomes clear.

Face areas are calculated by summing over a set of triangular areas: the area of face 2 in the example of Fig. 1 and Table I is given by summing the areas (calculated trivially from the vertex coordinates) of the three triangles (125), (154), and (143) "hinged" at the arbitrary "origin" vertex 1. The volume subtended by the face is found from the sum of the volumes of the three tetrahedra formed with the above three triangles as bases, the centre 1 of the polyhedron serving as the fourth tetrahedron

vertex. The volume of the total polyhedron follows by summation over volumes subtended by all faces. A complete topological and quantitative specification of the polyhedron is thus obtained.

3. PROGRAMMING CONSIDERATIONS

In applying this basically simple approach, a subset of the whole assembly of N centres must be selected for the solutions of Eq. (1) for each polyhedron. Otherwise the work done would increase as $\binom{N}{4}$, or $O(N^4)$ for large N , which would be impracticable for most assemblies. Thus a subset of N_c centres within a "combination sphere" radius R_c , centred on the centre whose polyhedron is to be elucidated, is extracted, and Eq. (1) solved for all possible combinations of three of this subset with the central centre. This cuts the work associated with vertex solution to $O(N_c^3)$ for each polyhedron, or $O(N_c^3 N)$ in all.

The choice of optimum R_c is made on the basis of results from trial runs. For the homogeneous dense single-component systems such as liquids and glasses that have been the major object of polyhedron studies to date, we take $R_c \sim 1.6D_o$, where D_o is the closest approach of two "atoms" in the molecular assembly. Referring to the radial distribution functions of such assemblies [4, 6] we find about $N_c = 16$ centres occur within a combination sphere so defined; this is $\sim \bar{F} + 2$, where \bar{F} is the average number of faces per polyhedron in such systems.

One further set of tests is required. Not every solution of Eq. (1) for all possible (jkl) from the subset N_c will be a valid vertex: only those closer to the centre i than to any other centre chosen from the complete assembly of N are acceptable. To speed up this checking procedure, a second subset of N_s centres within a "search sphere" radius R_s from the centre i is set up during the identification of the combination subset N_c . In principle for all possibly significant members of N_s , $R_s \sim 2R_c$; in practice, vertices rarely occur further than $0.9D_o$ from centre i , and $R_s \sim 2.0D_o$ is generally used. This corresponds to about $N_s = 45$ checks that need to be made for each trial vertex, although only a fraction of these need be made before an invalid vertex is detected.

The self-checking nature of the vertex-elucidation procedure prevents any errors arising from inadequate choice of R_c or R_s . The method used to order vertices round a face fails to close if N_c is not sufficient; when this occurs, R_c and R_s are incremented slightly and the procedure restarted. Moreover, for the checking of possible vertices against the subset N_s to be exhaustive, r (Eq. (1)) must be less than $R_s/2$. In heterogeneous systems this condition may occasionally fail to be fulfilled, in which case R_s is increased automatically as necessary. In fact the program can be modified to optimise the choice of R_s automatically as the polyhedron elucidation proceeds.

Similarly, a polyhedron at the surface may fail to close. Alternatively a surface polyhedron may be found which is distorted from what it would be were the centre embedded within a large equivalent assembly. These problems are avoided by placing additional points at the surface (e.g., by placing further model atoms at tetrahedral

surface sites), and rejecting those polyhedra which involve these points. Such procedures are clearly unnecessary when examining assemblies with periodic boundaries.

It should be emphasised that R_c and R_s are quantities whose optimum values are fixed empirically for a given system to reduce the computing time required. Our theoretical understanding of non-crystalline dense packings of hard spheres, for example, is insufficient to allow us to fix either search or combination radius to that (maximum) value required to ensure all polyhedron vertices will be elucidated using a single value of each of these quantities. It would not be particularly useful to do so even if we could; the systems we examine by this technique show local variations so that using just that value of, say, R_c which would ensure no further checks on vertices would be required would of necessity mean that the subset N_c would be too large for most of the centres in the assembly. Provided that an initially low value of R_c is recognised in the program without giving rise to any errors in the resultant polyhedron, an empirical, variable choice of this radius is to be preferred over a rigidly-fixed, and necessarily inefficient one.

The procedure described in reference [29] runs at a reasonable speed without the need to make such empirical choices. Savings in both the time and storage requirements of that procedure may be possible through the use of similar devices.

4. TIME AND STORAGE REQUIREMENTS

We can estimate the expense of the major computing operations, using as a basic time unit that required to test whether two given points in three-dimensional space are closer than a given distance. Each solution of Eq. (1) requires about ten such units, resulting in the factor of ten in operation (b) below.

(a) Selection of "combination" and "search" subsets N_c and N_s for the whole assembly of N centres is proportional to N^2 .

(b) Solving Eq. (1) for possible vertices using subset N_c . This is $\sim |N_c|/3!(N_c - 3)! \times 10$, or $\sim 3NN_c$ for all N polyhedra.

(c) Checking that trial vertices from (b) are no closer to any member of the N_s subset. Although $N_c^3/3$ trial vertices must be checked, few will require checking against more than a fraction of N_s before they are rejected. A reasonable time estimate is $\sim N_c^3 N_s/6$ per polyhedron.

(d) Vertex sorting and area and volume calculations. This work is minor compared with (a)–(c) and increases linearly with N .

When typical values of N_c and N_s are considered, these time estimates are comparable with those estimated by Brostow et al [29] for their procedure, whose major contributions increase as $O(N^2 \log N)$ and $O(N_c^3 N)$ (compare operations (a), and (b) + (c) respectively). For well-behaved systems, taking $N_c \sim 16$, $N_s \sim 45$, operations (b) and (c) are the major time-consuming steps, with (c) involving 2–3 times the work of (b). For $N \gtrsim 4000$, the selection of subsets (operation (a)) begins

to be comparable. For less well-behaved heterogeneous systems such as proteins and nucleic acids, N_c may reach 25, with a corresponding increase in N_s ; as the work done depends more critically on polyhedron size, this considerably increases the time consumed over the well-behaved dense liquids and glasses by a factor of about 2–6.

Further modifications are possible [30] to reduce the time consumption of steps (a) and (b) for sufficiently large numbers of points in the array. Selection of subsets N_c and N_s can be made from subsets into which the whole array is earlier partitioned, for example by using standard boxing procedures. Operation (b), which consumes the major portion of the central processor time, could be speeded up by further partitioning of the subset N_c into smaller subsets $N_{c'}$, each corresponding to every possible of j within N_c . We then need to solve Eq. (1) for k and l values chosen from the reduced set $N_{c'}$. Additional searching time would be required to set up the subsets $N_{c'}$ for each chosen centre j (i given); we estimate savings of 30% could accrue from choosing members of $N_{c'}$ to be those (kl) within the hemisphere with vector ij parallel to the axis of revolution of the hemisphere, radius R_c .

Concerning storage requirements, because of our applications to large and heterogeneous assemblies, we have preferred to buy low storage with increased time consumption. With our procedure, each vertex is elucidated independently four times, once for each of the (generally) four polyhedra in which it occurs. To store each vertex in terms of its four indices would, for a well-behaved assembly with an average polyhedron of, say, fifteen faces, and hence 26 vertices, require $26 \times 4N/4$ integer storage locations. For an assembly of 5,000 centres using half-word integers on an IBM370 we would thus require ~ 260 kb; for more heterogeneous systems we would expect around 46 vertices per polyhedron, resulting in a prohibitive fast core penalty of about 460kb. Additional computing necessary for searching existing vertices would cut down significantly on the time saved by elucidating each polyhedron once only. Storing faces only would require half the above amount of store, but with duplicating computing to derive vertices therefrom.

This program, written in standard FORTRAN IV, has been used effectively since 1966 on a variety of assemblies from crystals, through simple dense homogeneous non-crystalline assemblies such as liquids and glasses, to the more heterogeneous proteins and nucleic acids. A two-dimensional version is also operative.

REFERENCES

1. J. D. BERNAL, *Proc. Roy. Soc. London Ser. A* **280** (1964), 299.
2. A. RAHMAN, *J. Chem. Phys.* **45** (1966), 2584.
3. J. D. BERNAL AND J. L. FINNEY, *Discuss. Faraday Soc.* **43** (1967), 62.
4. J. L. FINNEY, *Proc. Roy. Soc. London Ser. A* **319** (1970), 479.
5. J. L. FINNEY, *Proc. Roy. Soc. London Ser. A* **319** (1970), 495.
6. J. L. FINNEY, Ph.D. thesis, Univ. of London, 1968.
7. D. H. EVERETT, *Discuss. Faraday Soc.* **43** (1967), 82.
8. R. COLLINS, *Proc. Phys. Soc.* **86** (1965), 199.
9. T. KIANG, *Z. Astrophys.* **64** (1966), 433.
10. J. L. MEIJERING, *Philips Res. Rep.* **8** (1953), 270.

11. J. L. FINNEY, *J. Phys. Paris* **36** (1975), Colloque C2 — 1.
12. J. L. FINNEY, *Mater. Sci. Eng.* **23** (1976), 199.
13. C. H. BENNETT AND J. L. FINNEY, in "Rapidly Quenched Metals" (N. J. Grant and B. C. Giessen, Eds.), p. 231, MIT Press, Cambridge, 1976.
14. J. A. BARKER, M. R. HOARE, AND J. L. FINNEY, *Nature* **257** (1975), 120.
15. J. L. FINNEY, in "The Structures of Non-crystalline Materials" (P. H. Gaskell, Ed.), p. 35, Society of Glass Technology, Taylor & Francis, London, 1977.
16. J. L. FINNEY, *Nature* **266** (1977), 309.
17. M. TANEMURA, Y. HIWATARI, H. MATSUDA, T. OGAWA, N. OGITA, AND A. UEDA, *Progr. Theor. Phys.* **58** (1977), 1079.
18. A. BONISSENT, J. L. FINNEY, AND B. MUTAFTSCHIEV, in "Proc 7th Intern. Vac. Congr. and 3rd Intern. Conf. Solid Surfaces, Vienna, 1977," p. 44.
19. A. BONISSENT, J. L. FINNEY, AND B. MUTAFTSCHIEV, in "4th European Crystallographic Meeting, Oxford, 1977," Vol. B, p. 651.
20. F. M. RICHARDS, *J. Mol. Biol.* **82** (1974), 1.
21. C. CHOTHIA, *Nature* **254** (1975), 304.
22. J. L. FINNEY, *J. Mol. Biol.* **96** (1975), 721.
23. F. M. RICHARD, *Ann. Rev. Biophys. Bioeng.* **6** (1977), 151.
24. J. L. FINNEY, *J. Mol. Biol.* **119** (1978), 415.
25. J. L. FINNEY, in "Water: A Comprehensive Treatise" (F. Franks, Ed.), Vol. 6, Plenum, New York, in press.
26. W. BROSTOW AND Y. SICOTTE, *Physica A* **80** (1975), 513.
27. A. L. MACKAY AND J. L. FINNEY, *Acta Crystallogr. Sect. A* **28** (S4) (1972), S110.
28. A. L. MACKAY, *J. Microscopy* **95** (1972), 217.
29. W. BROSTOW, J. P. DUSSAULT, AND B. L. FOX, *J. Computational Physics*, **29** (1978), 81.
30. I am indebted to one of the reviewers for helpful comments from which these points originated.

RECEIVED: May 15, 1978; REVISED: October 18, 1978

J. L. FINNEY
Department of Crystallography
Birbeck College, University of London
Malet Street, London WC1E 7HX, England