FU-WEN LING & STARKE, E. A. JR (1971). *Acta Met.* **19**, 759–768.
GOLDAK, J., BARRETT, C. S., INNES, D. & YOUDELIS, W. (1966). *J. Chem. Phys.* **44**, 3323.
JANSSEN, H. & BOON, M. (1967). *Theory of Finite Groups. Applications in Physics*. Amsterdam: North-Holland.
MELVIN, M. A. (1956). *Rev. Mod. Phys.* **28**, 18–44.
OKAMOTO, P. R. & THOMAS, G. (1971). *Acta Met.* **19**, 825–841.
RUEDL, E. & AMELINCKX, S. (1969). *Mater. Res. Bull.* **4**, 361–368.
RUEDL, E. & AMELINCKX, S. (1971). *Cryst. Latt. Defects*, **2**, 247–249.
RUEDL, E., DELAVIGNETTE, P. & AMELINCKX, S. (1968). *Phys. Stat. Sol.* **28**, 305–328.
SNYKERS, M., DELAVIGNETTE, P. & AMELINCKX, S. (1972). *Mater. Res. Bull.* **7**, 831–840.
STOEMENOS, J. & VINCENT, R. (1972). *Phys. Stat. Sol.* (*a*), **11**, 545–558.
VAN LANDUYT, J. & AMELINCKX, S. (1969). *Phys. Stat Sol.* **31**, 589–600.
VAN LANDUYT, J. & AMELINCKX, S. (1972). *Mater. Res. Bull.* **7**, 71–79.
VAN LANDUYT, J., REMAUT, G. & AMELINCKX, S. (1970). *Phys. Stat. Sol.* **41**, 271–289.
VAN TENDELOO, G. & AMELINCKX, S. (1973). *Mater. Res. Bull.* **8**, 721–732.
VAN TENDELOO, G., DELAVIGNETTE, P., GEVERS, R. & AMELINCKX, S. (1973). *Phys. Stat. Sol.* (*a*), **18**, 85–97.

*Acta Cryst.* (1974). A**30**, 440

# Generalized Structural Geometry

BY ALAN L. MACKAY

*Department of Crystallography, Birkbeck College (University of London), Malet Street, London WC1E 7HX, England*

(*Received* 11 *October* 1973; *accepted* 2 *January* 1974)

Arbitrary combinations of bond lengths, bond angles and torsion angles can be used as generalized coordinates for describing molecular models. It is shown how these and the conventional Cartesian or fractional unit-cell coordinates can be interconverted. Algorithms are also given for the geometrical analysis of rigid structures of links joined by flexible connectors (where only bond lengths are specified). Properties of the connectivity matrix, as an alternative complete description of a structure, are developed. Several applications of the above procedures are described.

## Introduction

Given a set of parameters sufficient to define a general, unsymmetrical geometrical structure, such as a polyhedron (convex or otherwise), a molecule or a framework, it is often necessary to calculate certain consequential geometrical parameters. This calculation is often very difficult or tedious when done by exact algebraic means, since the solution of several simultaneous quadratic equations is usually necessary. For calculations of the energy and other parameters depending on pair interactions, for example, all interatomic distances must be found. Problems similar to those of molecular architecture occur also in real architecture, particularly with structures like the geodesic domes elaborated by Buckminster Fuller.

Modern computational facilities, especially those providing on-line access by time sharing, are now widely available. The procedures to be described show how geometrical models, very similar to the widely used ball-and-spoke models, but of indefinitely high accuracy, can be realized with a general computer. The programs have been written in BASIC and, since a program itself is very easily changed while using it from a teletypewriter keyboard, elaborate provisions for all eventualities are not necessary. The programs (on paper tape) are available from the author.

## A general geometrical program

The primary or intrinsic parameters of an array of $N$ points are the bond distances $d_{ij}$ between pairs of points, the bond angles $\theta_{ijk}$ defined by triplets of points and running from 0 to 180° and the torsion angles (or dihedral angles) $\varphi_{ijkl}$ which each require four points for identification and which run from $-180$ to 180°, a sense of rotation being defined in terms of a right-hand screw (Appendix). The secondary or derived parameters of the points are their coordinates $x_i, y_i, z_i$ with respect either to orthogonal Cartesian axes or to crystallographic axes. The derived coordinates change with the axes and are not invariants of the structure. Following the Erlangen Programme of Klein (1872) wherever possible, quantities which are invariants of the structure and thus of physical significance are used. In each case $3N$-6 parameters are needed to describe a structure of $N$ points. Six further parameters clamp the grouping in a coordinate system.

Just as the solution of triangles is fundamental

to plane geometry, so solid geometry requires the solution of tetrahedra. Often no single triangle of the four making any tetrahedron can be solved directly from the data provided, and simultaneous equations must be set up. In a tetrahedron six properly-chosen parameters are sufficient for its definition, but any of six distances, twelve bond angles and six dihedral angles may be required. Euclid's Elements give no guidance on irregular tetrahedra and, although formulae were developed in the eighteenth century, it is simpler to use numerical methods. A program to solve tetrahedra is desirable but, as there are many structures which cannot be divided into tetrahedra, a more general approach has been followed.

A general program (*MOLEC*4) has been written in BASIC to solve a wide variety of problems of this category. In outline its operation is as follows:

Initial data are supplied. These are:

(*a*) Unit-cell parameters if fractional cell coordinates are to be used; otherwise $1,1,1,90°,90°,90°$ are assumed as cell dimensions.

(*b*) The number of points ($N$) in the structure.

(*c*) The $x,y,z$ coordinates of the first three points, which are to be taken as fixed. (This is not the only way in which a structure may be clamped in a framework of coordinates and it fixes nine parameters, instead of the six parameters which can be chosen arbitrarily. However, if other methods of fixing the structure are required, then extra dummy points, not part of the structure, can be introduced as a device for specifying the particular constraints, which may be angles relating the structure to the axes. In this way the proper number of six parameters can be kept fixed while the others are refined.)

(*d*) Rough $x,y,z$ coordinates for each of the other points. (These are the quantities which are to be refined, so that if they are initially close to the correct values refinement will be quicker. Refinement will nevertheless usually occur even if the initial values are far out. The initial coordinates are really only needed to guide the refinement into the desired configuration rather than let it go into another minimum, in the multidimensional coordinate space, which may also satisfy the requirements.)

(*e*) The number of interpoint distances $d_{ij}$, the distances themselves, their estimated standard deviations and their reference numbers $i$ and $j$.

(*f*) Similar data for the bond angles $\theta_{ijk}$ and the torsion angles $\varphi_{ijkl}$. (It is usual to specify standard deviations of 0·001 Å for distances, 0·1° for bond angles and 1° for torsion angles, to resemble the properties of a ball-and-spoke model, but this is only used to determine the weights in the least-squares refinement and does not affect the overall accuracy.)

The program operates by calculating the values of the given parameters from the initial coordinates and finding the discrepancies. The weighted sum of the squares of the discrepancies is minimized. Each coordinate $x_i,y_i,z_i$ is then varied in turn and the discrepancies are recalculated to give the differential coefficients with which to form the observational equations. The weighted observational equations are then combined to give the normal equations which are then solved. Since BASIC has exceptionally concise matrix statements, this process is simple. The corrections to the initial positions $x_i,y_i,z_i$ are applied but, as the problem is non-linear, the procedure must be iterated. The corrections applied to give positions for the next cycle are severely damped so that, for typical molecular structures, the maximum displacement allowed is about 1·7 Å, one typical bond distance. This is done by using the arctangent function of $x$, which is near $x$ when the latter is small compared with one, but which does not exceed $\pi/2$ however big $x$ may be. When the corrections to the coordinates do not exceed a preset amount (usually $10^{-6}$ Å) refinement ceases and the initial and final coordinates are printed out, together with the prescribed values of the parameters and their final values. If the problem is exactly determined by the initial data, then an exact fit with the data is obtained; if there are more data than necessary, then a least-squares fit is obtained and if the problem is underdetermined, the refinement fails. If required for the least-squares case, standard deviations could be calculated from the normal matrix, but this is usually unnecessary and is omitted.

Normally the corrections to the coordinates are printed out at each cycle to enable the process of convergence to be followed. If the refinement does not take place in a few cycles then the situation demands inspection which the running of the programme on-line permits.

If, for example, a helix is specified by the coordinates of the first group together with the torsion angles at which successive groups should be added, then a later residue cannot be refined into position until its predecessor is settled. It must, therefore, be prevented from running away before its turn comes. The mathematical process thus clearly represents the physical procedure of model building; if a physically impossible procedure is demanded then, of course, the program will fail. At the end of the program any desired parameters can be calculated from the coordinates which have been found. Run-time input is used for this.

## Applications

Some typical applications of the procedure are the following:-

(*a*) Obtaining the $x,y,z$ coordinates of the atoms in a helix given the distances and angles relating one group to another.

(*b*) Obtaining the whole set of fractional coordinates in a crystal-structure analysis when part of the molecule has been recognized.

(*c*) Finding the stresses in an engineering structure by calculating how the whole set of parameters varies

when one is altered slightly (Maxwell's method of virtual work) (Timoshenko & Young, 1965).

(*d*) The problem readily solves the general problem discussed by Drazin & Otte (1963) and by Fong (1973) which is to assign indices to a crystal face, given that three other planes intersect it to give trace lines in the surface, the angles between which can be measured. In origin the problem is a particular metallurgical one and relates to the traces of the {111} slip planes which may appear on an etched surface of a cubic crystal. This reduces to the solution of a tetrahedron (Fig. 1) where six parameters are given. Angles *BAC*, *DAC*, *BAD*, *BDC*, *BCD* and the length *AD* are known. It is required to find the sides *BA* and *CA*. No single triangle can be solved by itself so that, as Fong's paper demonstrates, the algebra is laborious.

(*e*) Given five of the six angles between four bonds meeting at a point, it is required to find the sixth angle. This application is discussed further below.

(*f*) A further problem, discussed below, is to find the three undefined interpoint distances (12, 34, 56) in an irregular octahedron when the lengths of the 12 edges are given (Fig. 2). This is one example of a minimally connected structure of rods, freely jointed, which contains no tetrahedra or other separately rigid subunits. If any bond is cut the rigidity of the structure disappears, and thus the length of each bond depends on the length of every other. If the energy of atoms assembled in this way is to be calculated, then these extra distances have to be found.

(*g*) Refining a molecular structure in real space to obtain the $x,y,z$ coordinates which best fit an *a priori* set of interatomic distances and angles.

## Simplifications

It has been found convenient to use two simplified programs, both in principle contained in *MOLEC*4, in an extended examination of polyhedra. In the version *ICOSA*, coordinates $x_i,y_i,z_i$ are refined when only a minimal set of distances is supplied [application (*f*) above]. Here the differentials, the variations of the discrepancies with changes in the coordinates, are calculated analytically, instead of altering each coordinate in turn and recalculating all parameters. All distances are given equal weights and no provision is made for the least-squares treatment of excess data. With this program, for example, the coordinates of 12 points forming an irregular icosahedron can be found, given the lengths of the 30 unequal edges.

With this it is convenient to use a separate program (*DIMEN*) which calculates specified distances, bond angles and torsion angles, given a set of coordinates.

## Non-redundant rod structures

We will now examine the particular class of rigid structures where only the bond (or interpoint) distances $d_{ij}$ are prescribed and the values of the bond angles and the torsion angles follow as consequences. This class comprises those non-redundant structures which can be made by joining rigid rods with flexible connectors. An octahedron of 12 rods is such a structure. Cauchy's theorem (Lyusternik, 1956) states that if every face of a convex polyhedron is rigid, then the polyhedron as a whole is rigid. A triangle is the only polygon which can be rigid without redundancy, and therefore the class is that of the deltahedra and more general triangulated structures.

Manipulations use only interpoint distances, which are most conveniently presented as a matrix.

## Properties of the metric matrix

We will examine the properties of the metric matrix, first for the usual three-dimensional case (*International Tables for X-ray Crystallography*, 1959, p. 60).

The metric matrix in three dimensions is important because it represents a tetrahedron, the solution of which, as we have seen, is the primary requirement in dealing with three-dimensional structures.

Three non-coplanar vectors **a**, **b** and **c** define a parallelepiped *OAB'CC'O'BA* (Fig. 3). *O* is the origin. The metric matrix $g_{ij}$ is independent of the particular axes to which the parallelpiped may be referred and has a number of important properties. It is defined as the array of the scalar products of the edges.

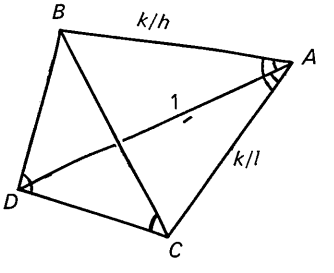


Fig. 1. The traces of the planes 0*kl*, *h*0*l* and *hk*0 on *hkl* intersect at measured angles *BDC* and *BCD*. Given the cell dimensions and these angles find *hkl*.

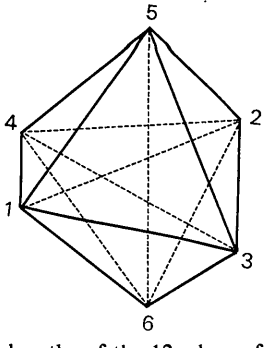


Fig. 2. Given the lengths of the 12 edges of an irregular octahedron 123456, find the distances 12, 34, 56.

$$[g_{ij}]=\begin{bmatrix}\mathbf{a}\cdot\mathbf{a} & \mathbf{a}\cdot\mathbf{b} & \mathbf{a}\cdot\mathbf{c}\\ \mathbf{b}\cdot\mathbf{a} & \mathbf{b}\cdot\mathbf{b} & \mathbf{b}\cdot\mathbf{c}\\ \mathbf{c}\cdot\mathbf{a} & \mathbf{c}\cdot\mathbf{b} & \mathbf{c}\cdot\mathbf{c}\end{bmatrix}.$$

Klein points out (p. 145) that 'The theory of invariants permits the systematic enumeration of all possible magnitudes and theorems, without exception, since it supplies the complete system of invariants and syzygies'.

The invariants of the metric matrix are:
The trace, $T=a^2+b^2+c^2=\frac{1}{4}$ (sum of the squares of the 12 edges of the parallelepiped). The second invariant,

$$S=\begin{vmatrix}b^2 & b\cdot c\\ c\cdot b & c^2\end{vmatrix}+\begin{vmatrix}a^2 & a\cdot b\\ b\cdot a & b^2\end{vmatrix}+\begin{vmatrix}a^2 & a\cdot c\\ c\cdot a & c^2\end{vmatrix}=\frac{1}{2}$$

(sum of the squares of the areas of the six faces of the parallelohedron). These three terms are each separately invariant, as are, of course, the lengths of the edges.

The determinant, $D=|g_{ij}|=$ the square of the volume of the parallelohedron.

The volume of the simplex $OABC$ is $\frac{1}{6}$ of that of the parallelepiped and in $N$ dimensions $1/N$ of the hypervolume of the corresponding parallelohedron. If the characteristic equation $|g_{ij}-\delta_{ij}\lambda_j|=0$ is set up, then these invariants are the coefficients.

If the three angles at $O$ are right angles, then we find the special relationship, an extension of Pythagoras' Theorem, that for a right-angled tetrahedron $OABC$, the square of the area of the hypotenuse $ABC$ (extending the usual meaning of the word) is equal to the sum of the squares of the areas of the other three sides. This is so because in three dimensions, as in two, the sum of the squares of the direction cosines is unity.

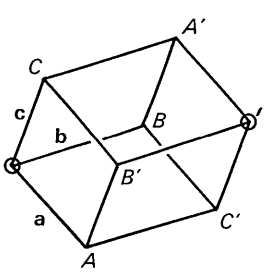


Fig. 3. A parallelepiped is defined by three vectors **a**, **b** and **c**.

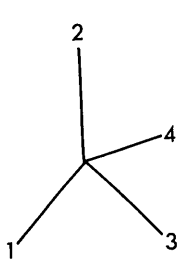


Fig. 4. Given five of the six angles between four bonds meeting at a point, find the sixth.

### The $N$-dimensional metric matrix

Similar relationships hold in higher numbers of dimensions. For example, in four dimensions four vectors $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3,\mathbf{a}_4$ define a simplex of five points (counting the origin) and a parallelohedron of 16 vertices. The metric matrix is $[g_{ij}]=[\mathbf{a}_i\cdot\mathbf{a}_j]$. Its invariants are:- Trace $=T=\frac{1}{8}$(sum of squares of 32 edge lengths). Second invariant $=S=\frac{1}{4}$ (sum of squares of areas of 24 faces). Third invariant $=U=\frac{1}{2}$ (sum of squares of volumes of 8 parallelepipeds). Determinant $=D=$ square of hypervolume of one 4-dimensional parallelohedron. (The Euler relation is satisfied since $16-32+24-8+1=1$.) If the simplex is of dimensionality less than four then the hypervolume vanishes. If, further, we take four unit vectors then the scalar products $\mathbf{a}_i\cdot\mathbf{a}_j$ represent the cosines of the angles $\theta_{ioj}$ (abbreviated to $\theta_{ij}$). Thus,

$$0=\begin{vmatrix}1 & \cos\theta_{12} & \cos\theta_{13} & \cos\theta_{14}\\ \cos\theta_{12} & 1 & \cos\theta_{23} & \cos\theta_{24}\\ \cos\theta_{13} & \cos\theta_{23} & 1 & \cos\theta_{34}\\ \cos\theta_{14} & \cos\theta_{24} & \cos\theta_{34} & 1\end{vmatrix}.$$

This is thus the relationship connecting the six angles between four bonds meeting at a point (Fig. 4). If five of these angles are known, then the sixth can be calculated, but not without some difficulty if it has to be done algebraically, although the determinant can be multiplied out to give a quadratic for one of the angles in terms of the others. It follows that the value of this determinant is a proper measure of the consistency of a set of six independently measured bond angles. The calculation is, in fact, best done by the computational algorithm given earlier (Mackay, 1973).

I am indebted to Professor J. D. Dunitz for pointing out this application.

In three dimensions the analogous expression is already familiar and if the determinant for the volume of a cell is multiplied out we obtain $V^2=a^2b^2c^2(1-\cos^2\theta_{12}-\cos^2\theta_{23}-\cos^2\theta_{31}+2\cos\theta_{12}\cdot\cos\theta_{23}\cdot\cos\theta_{31})$. The condition that the expression in brackets should be zero is thus $\theta_{12}+\theta_{23}+\theta_{31}=360°$.

In general, the metric matrix in $N$ dimensions is of rank $N$, but, if the parallelohedron or simplex (of $N+1$ points) which is described is degenerate and is only $N$-dimensional, then the $N\times N$ metric matrix is only of rank $M$. This means that all minors of rank greater than $M$ will be zero. This can be used, as Klein indicates, to provide a number of powerful relationships for solving the geometry of rod structures. We may take two examples.

### The pentacle problem

In three dimensions, given the distance of a point 1 from three reference points 3, 4, 5 and the distances

of a second point 2 from the same three points, find the distance $d_{12}$. Two solutions are possible; 1 and 2 on the same side of 345 and 1 and 2 on opposite sides of 345. The problem is a basic one in calculating all distances in a structure and is the three-dimensional analogue of triangulation (Fig. 5). Since the figure 12345 is three-dimensional, its four-dimensional hypervolume will be zero, thus:

$$\begin{vmatrix} 0 & d_{12}^2 & d_{13}^2 & d_{14}^2 & d_{15}^2 & 1 \\ d_{21}^2 & 0 & d_{23}^2 & d_{24}^2 & d_{25}^2 & 1 \\ d_{31}^2 & d_{32}^2 & 0 & d_{34}^2 & d_{35}^2 & 1 \\ d_{41}^2 & d_{42}^2 & d_{43}^2 & 0 & d_{45}^2 & 1 \\ d_{51}^2 & d_{52}^2 & d_{53}^2 & d_{54}^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{vmatrix} = 0.$$

This determinant relating the distances $d_{ij}$ between the points $i$ and $j$ is symmetrical and is a quadratic in $d_{12}^2$. It can be solved by the algorithm mentioned which involves iteration, or can be solved directly from the quadratic $Ad_{12}^4+Bd_{12}^2+C=0$ where the coefficients are:

$$A=\begin{vmatrix} 0 & d_{34}^2 & d_{35}^2 & 1 \\ d_{43}^2 & 0 & d_{45}^2 & 1 \\ d_{53}^2 & d_{54}^2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{vmatrix},$$

which is $16\times$(area of the triangle 345)$^2$,

$$B=2\begin{vmatrix} 0 & d_{13}^2 & d_{14}^2 & d_{15}^2 & 1 \\ d_{32}^2 & 0 & d_{34}^2 & d_{35}^2 & 1 \\ d_{42}^2 & d_{43}^2 & 0 & d_{45}^2 & 1 \\ d_{52}^2 & d_{53}^2 & d_{54}^2 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}$$

and $C$ is equal to the original determinant in which $d_{12}^2$ is set equal to zero.

Since $M=Ax^2+Bx+C$, on differentiation we have
$M'=2Ax+B$ so that $B=(M')_{x=0}$
and $M''=2A$.

### Solution of the octahedral frame (Fig. 2)

The distances $d_{ij}$ between the six points are used to form a symmetrical $6\times6$ matrix $[d_{ij}^2]$. This we call the connectivity matrix. This is then expanded to give the expression for the five-dimensional hypervolume, a $7\times7$ determinant exactly analogous to the one used in the previous section. By subtracting the first row and the first column from the others, since $\mathbf{d}_{12}\,.\,\mathbf{d}_{13}=\frac{1}{2}(d_{13}^2+d_{12}^2-d_{23}^2)$ we obtain the metric matrix with origin at the point 1.

In the case of the octahedron the terms $d_{12}$, $d_{34}$ and $d_{56}$ are to be found, all the 12 others being known. Since the figure is three-dimensional, the four-dimensional hypervolumes, which are the minors of the above determinant are zero. There are three different minors of this type, so that we have three equations for three unknowns. An iterative computer method of solution is necessary. However, the program *ICOSA*, discussed above, is in fact a simpler method. The example demonstrates the situation when there are no constituent tetrahedra in a framework.

### Factorization of the metric matrix

In pursuit of the aim of using invariants (such as distances) instead of coordinates in the description of structures, it is necessary to be able to pass easily from one system to the other. It is clearly easy to calculate distances, bond angles and torsion angles when given the coordinates, and this is done by the program (*DIMEN*) mentioned above. To go in the opposite direction requires the following procedure.

The $(N-1)\times(N-1)$ metric matrix formed as above from the squares of the distances between $N$ points in $M$ dimensions is symmetrical and can thus be factorized by the Choleski process (Rollett, 1965) into the product of a lower triangular matrix and its transpose. A program (*MATFAC*) has been written in BASIC to do this.

We will show that, if $M=3$ and the structure consists of $N$ points (for example six points forming an octahedron) the resulting lower triangular matrix gives the $x_i,y_i,z_i$ coordinates of the points with respect to orthogonal Cartesian axes where the point 1 is taken as origin, the point 2 lies on the $x$ axis, the point 3 lies in the $xy$ plane and the point 4 has a positive $z$ coordinate (Fig. 6). That is, the tetrahedron 1234 is held in the orthogonal axes as if fixed by a 'kinematic clamp (Elliott & Dickson, 1951).

Taking first only four points 1234, we have the expression (*International Tables for X-ray Crystallography* 1959, p.44) for the volume of the tetrahedron formed:-

$$6V=\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}$$

where $x_1=y_1=z_1=y_2=z_2=z_3=0$.

When this expression is multiplied by its transpose we have:

$$36V^2=[g_{ij}]=[\mathbf{a}_i\,.\,\mathbf{a}_j]$$

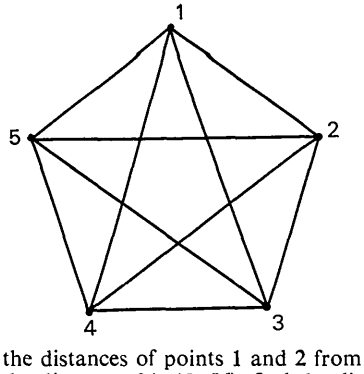


Fig. 5. Given the distances of points 1 and 2 from points 3, 4 and 5 (and the distances 34, 45, 56), find the distance 12.

which is the metric matrix since

$$\mathbf{a}_i \cdot \mathbf{a}_j = x_i x_j + y_i y_j + z_i z_j .$$

It is clear that this procedure can be carried out for any number of dimensions. If the configuration of $N$ points is degenerate and does not represent a simplex in $N$–1-dimensional space, but is fully described in three dimensions, then the coordinates relating to the further dimensions will be zero. The procedure outlined above thus derives the dimensionality and coordinates for a configuration of points from a table of all their mutual distances. This table must be a full one of all $\frac{1}{2}(N-1)N$ distances and not just the $3N-6$ distances which may be the minimum necessary. This gives rise to the necessity for the calculation of the extra distances which is described above.

If the distances are experimentally derived for a three-dimensional array of more than four points, then there will be residual errors which appear as coordinates in higher dimensions. The nature of the Choleski process is such that these residuals accumulate towards the lower right corner of the resultant matrix. They give a measure of the self-consistency of the interpoint distances, just as the hypervolume gives a measure of the consistency of the six angles between the four bonds.

### The connectivity matrix

The complete table of the interatomic distances in a molecule, which we regard as the fundamental variety of connectivity matrix, has further properties.

### Inertia

For example, the principal moments of inertia of a molecule composed of atoms of mass $m_i$ with interatomic distances $d_{ij}$ can be calculated without further data.

A theorem due to Lagrange (Lamb, 1911) demonstrates that the trace $T$ of the inertia tensor is given by $T$

$$T = \sum_i \sum_j m_i m_j (d_{ij})^2 / \sum_a m_i$$

(summed over all pairs).

It must be possible to find similar expressions for $S$, the second invariant of the inertia tensor and for $D$, its determinant, but the algebra is almost prohibitive. $S$ involves sums over vector and scalar products of pairs of $d_{ij}$ terms and $D$ requires sums over triple products. However, such expressions are made unnecessary by the procedure, described above, of factorizing the distance matrix and thus reducing the points to orthogonal coordinates from which the moments of inertia are readily calculated. The principal moments of inertia $I_1, I_2, I_3$ can be found by diagonalizing the matrix just calculated or by the equivalent procedure of finding the invariants which are $T = I_1 + I_2 + I_3$, $S = I_1 I_2 + I_2 I_3 + I_3 I_1$, $D = I_1 I_2 I_3$ and solving $\lambda^3 - T\lambda^2 + S\lambda - D = 0$ using the standard algorithm for cubic equations.

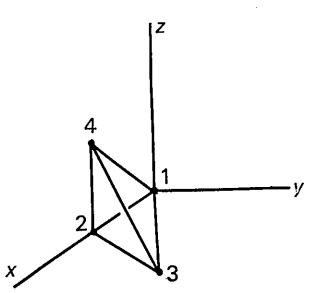


Fig. 6. Six parameters $(x_1, y_1, z_1, y_2, z_2, z_3)$ out of the $3N$ describing a configuration of $N$ points are arbitrarily made zero leaving the proper number of $3N-6$.

### Adjacency matrix

The matrix $[D]$ with elements $d_{ij}^2$ is readily converted to become an adjacency matrix $[C]$ by examining each distance $d_{ij}$ and replacing it by 1 if the points $i$ and $j$ are deemed to be connected and by 0 if they are not. The terms $C_{ii}$ are made zero.

The adjacency matrix $[C]$ can be used in examining the topology of a molecule. It is the basis of various methods of indexing molecular structure and accounts have been published as to how the adjacency matrix can be converted to the Wiswesser Line Notation or to the Chemical Abstracts Service line-by-line system or to the I.C.I. Crossbow indexing system. These systems can also be interconverted (Campey, Hyde & Jackson, 1970).

If the adjacency matrix $[C]$ is raised to the power $m$, then the entries represent the number of paths of length $m$ steps from point $i$ to point $j$. This may include many duplications, since, if $i$ is connected to $j$, then there are also $C_i + C_j$ paths of length three steps, where $C_i$ is the coordination number of $i$ and $C_j$ that of $j$. The sum of each row and column is the coordination number:

$$C_i = \sum_j C_{ij} = \text{coordination number of atom } i.$$

A true count of paths of length (for example) five steps where none is retraced is best achieved either by taking the sum $|\varepsilon_{ijklmn}| C_{ij} C_{jk} C_{kl} C_{lm} C_{mn}$ where $\varepsilon_{ijklmn}$ is the permutation tensor having the value 1 if $i,j,k,l,m,n$ are in an even permutation, $-1$ if they are an odd permutation and 0 if any two are equal. A subroutine to generate those values of the permutation tensor which are not zero can be used. For large numbers of steps this strategy is better than scanning all values and excluding those which are zero, which is more convenient for six or fewer steps.

All questions of topology and connectivity can be answered using the adjacency matrix.

### Other expressions

A number of other matrix expressions can be calculated from the distance matrix for special purposes

or can be expressed in this form. For example, various expressions for the mutual potential or mutual energy of atom pairs (for pairwise interactions) can be written instead of the distances. This shows how the molecule is subdivided into groups where interactions between atoms of the groups are strong and between atoms of different groups are weak. The matrix can be rearranged to be nearly decomposable.

The solid angle which one atom subtends at the face of the Voronoi polyhedron which it shares with another atom can be used as the entry in the matrix. This gives a picture of the atomic coordination, like that due to Pauling for the distribution of bond strength, but quantitative and susceptible of further development (Mackay, 1972).

## Summary

Table 1 shows how the various matrix representations of a molecular structure are interrelated.

Since this paper was submitted, Professor J.D.Dunitz has drawn my attention to a book, *Géométrie de Position* by L.N.M.Carnot (1803) in which the author formulated exactly the same problem, namely: 'In any system whatever of straight lines, in the same plane or not, being given certain of their lengths, or of the angles between them, or of the angles between planes containing them, in number sufficient to determine the figure, find the remainder of these parameters'. The problems which I have discussed relating to the general tetrahedron, the pentacle, the octahedron, the six angles between four bonds meeting at a point and the radius of a sphere tangent to four others (Mackay, 1973) are also solved in principle, although neither determinants, ideas of more than three dimensions nor computers were available at the time. Considerable progress was made towards the solution of the general problem posed and many results of the eighteenth century are collected, although usually without attribution.

## APPENDIX

### The description of molecules as linkwork computation of $d_{ij}$, $\theta_{ijk}$, $\varphi_{ijkl}$

$d_{ij}$ is the positive distance between point $i$ and point $j$. $\theta_{ijk}$ is the bond angle between the link $ij$ and the link $jk$ ($0 \leq \theta_{ijk} \leq 180°$). The torsion angle $\varphi_{ijkl}$ (Fig. 7) is the dihedral angle between the plane containing $i, j$ and $k$ and that containing $j$, $k$ and $l$ ($-180° \leq \varphi_{ijkl} < 180°$). In defining $\varphi$ a sense of rotation is required; this can be taken as positive for a right-hand screw motion from $ij$ to $kl$ about $jk$. The sign is provided by the sign of the volume of the tetrahedron $i, j, k, l$ which is given by the triple product of $\mathbf{r}_{ij}$, $\mathbf{r}_{jk}$ and $\mathbf{r}_{kl}$. If all four points are coplanar and the sign of the volume is undefined, then if $i$ and $l$ are on the same side of $j$, $k$, $\varphi = 0$ and if on opposite sides, $\varphi = 180°$. With

Table 1. *Varieties of the connectivity matrix for N points*

$$[D] = \begin{bmatrix} d_{11}^2 & d_{12}^2 & . \\ d_{21}^2 & d_{22}^2 & . \\ . & . & . \end{bmatrix} \longleftrightarrow \begin{vmatrix} 0 & d_{12}^2 & d_{13}^2 \ldots 1 \\ d_{21}^2 & 0 & d_{23}^2 \ldots 1 \\ . & . & . \ \ldots 1 \\ 1 & 1 & 1 \ \ldots 0 \end{vmatrix} = (-1)N_2N-1[(N-1)!]^2V^2$$

$d_{IJ}^2$ matrix $(N \times N)$

Determinantal expression for $(N-1)-$dimensional hypervolume. $(N+1) \times (N+1)$.

$$[g_{IJ}] = \begin{bmatrix} \mathbf{a}_1^2 & \mathbf{a}_2^2 . \mathbf{a}_2 & \mathbf{a}_1 . \mathbf{a}_3 \ldots \\ \mathbf{a}_2 . \mathbf{a}_1 & \mathbf{a}_2^2 \ldots & \\ \ldots\ldots\ldots & & \end{bmatrix} \text{ factorize } \rightleftharpoons \text{ multiply } \begin{bmatrix} 0 & 0 & 0 & & & \\ x_1 & 0 & 0 & & & \\ x_2 & y_2 & 0 & & & \\ x_3 & y_3 & z_3 & 0 & & \\ x_4 & y_4 & z_4 & 0 & 0 & \\ x_5 & y_5 & z_5 & 0 & 0 & 0 \\ . & . & . & 0 & . & . \end{bmatrix}$$

Metric matrix $(N-1) \times (N-1)$

Matrix of orthogonal coordinates
Lower triangular matrix $N \times N$.

Various functions of $d_{IJ}$

Invert

$[g_{IJ}]$
Inverse metric matrix.

Invert

$$\rightarrow \begin{bmatrix} 0 & 1 & 1 & . & . \\ 1 & 0 & 1 & . & . \\ 1 & 1 & 0 & . & . \\ . & . & . & . & . \end{bmatrix} = [C]$$

Adjacency matrix $(N \times N)$

Note:
$g_{IJ} = \mathbf{a}_I . \mathbf{a}_J = \frac{1}{2}(d_{I0}^2 + d_{J0}^2 - d_{I9}^2)$
$d_{IJ}^2 = (\mathbf{a}_I - \mathbf{a}_J)^2 = g_{II} + g_{JJ} - 2g_{IJ}$

→ $[\Omega_{IJ}]$
Voronoi solid-angle matrix.

→ $[E_{IJ}]$
Various expressions for energy from pairwise interactions.

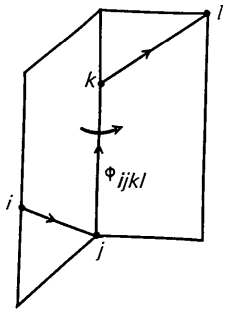


Fig. 7. The definition of the torsion or dihedral angle $\varphi_{ijkl}$.

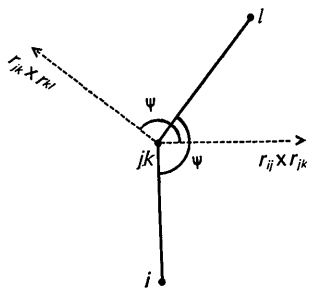


Fig. 8. Fig. 7 viewed along *kj*.

right-handed axes and the sequence $i,j,k,l, 0<\varphi_{ijkl}<180°$ corresponds to a negative volume. The convention is that of Klyne & Prelog (1960).

The angle between any two lines, even if they do not intersect, is also defined and is the angle at which they would intersect if they were translated so that they had a common point.

$d_{ij}$, $\varphi_{ijk}$ and $\varphi_{ijkl}$ can be calculated as follows, assuming that if necessary the crystallographic coordinates are first converted to Cartesian coordinates $x_i, y_i, z_i$. The vector from the origin $O$ to the point $P$ is denoted by $\mathbf{r}_p$.

(*a*) Thus, $d_{ij}=|\mathbf{r}_{ij}|=|\mathbf{r}_j-\mathbf{r}_i|=[(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2]^{1/2}$.

(*b*) The scalar product of two vectors **a** and **b** is

$\mathbf{a}\,.\,\mathbf{b}=|\mathbf{a}|\,|\mathbf{b}|\cos\theta_{ab}=a_xb_x+a_yb_y+a_zb_z$ so that

$$\theta_{ijk}=\arccos\left(\frac{\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk}}{|\mathbf{r}_{ij}|\,|\mathbf{r}_{jk}|}\right).$$

(*c*) Consider the projection of the links $i,j,k,l$ along the link $jk$ (Fig. 8).

Then $\varphi_{ijkl}=\psi$ and

$$\cos\psi=\frac{(\mathbf{r}_{ij}\times\mathbf{r}_{jk})\,.\,(\mathbf{r}_{jk}\times\mathbf{r}_{kl})}{|\mathbf{r}_{ij}\times\mathbf{r}_{jk}|\,|\mathbf{r}_{jk}\times\mathbf{r}_{kl}|}.$$

Using Lagrange's identity the numerator becomes:

$$(\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk})\,(\mathbf{r}_{jk}\,.\,\mathbf{r}_{kl})-(\mathbf{r}_{jk}\,.\,\mathbf{r}_{jk})\,(\mathbf{r}_{ij}\,.\,\mathbf{r}_{kl});$$

and since

$$|\mathbf{a}|\,|\mathbf{b}|=|\mathbf{a}\times\mathbf{b}|/\sin\theta_{ab}$$

the denominator equals

$$(\mathbf{r}_{ij}\times\mathbf{r}_{jk})\times(\mathbf{r}_{jk}\times\mathbf{r}_{kl})/\sin\psi\,.$$

Applying the expansion of the quadruple vector product, the denominator is then seen to be $(\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk}\times\mathbf{r}_{kl})\,(\mathbf{r}_{jk})$ so that after division,

$$\tan\psi=\frac{(\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk}\times\mathbf{r}_{kl})|\mathbf{r}_{jk}|}{(\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk})\,(\mathbf{r}_{jk}\,.\,\mathbf{r}_{kl})-\mathbf{r}_{jk}^2(\mathbf{r}_{ij}\,.\,\mathbf{r}_{kl})}.$$

This procedure is necessary because BASIC provides only arctangent as an inverse trigonometric function ATN, which is returned as an angle between 0 and 180°. We use the term $(\mathbf{r}_{ij}\,.\,\mathbf{r}_{jk}\times\mathbf{r}_{kl})$ which is the volume of the tetrahedron $i,j,k,l$ and if this is positive, subtract 180° from $\psi$.

## References


CAMPEY, L. H., HYDE, E. & JACKSON, A. R. H. (1970). *Chem. Brit.* **6**(10), 427–430.

CARNOT, L. N. M. (1803). *Géométrie de Position.* Paris.

DRAZIN, M. P. & OTTE, H. M. (1963). *Phys. Stat. Sol.* **3**, 824–858.

ELLIOTT, A. & DICKSON, J. H. (1951). *Laboratory Instruments*, p. 94. London: Chapman & Hall.

FONG, H. S. (1973). *Acta Cryst.* A**29**, 176–182.

*International Tables for X-ray Crystallography.* (1959). Birmingham: Kynoch Press.

KLEIN, F. (1872). *Elementary Mathematics: Geometry*, Reprinted 1939. New York: Dover Press.

KLYNE, W. & PRELOG, V., (1960). *Experientia*, **16**, 521–523.

LAMB, H. (1911). *Encyclopedia Britannica* (11th ed.). **17**, 972.

LYUSTERNIK, L. A. (1956). *Convex Figures and Polyhedra*, Reprinted 1963. New York: Dover Press.

MACKAY, A. L. (1972). *J. Microsc.* **95**,(2), 217–227.

MACKAY, A. L. (1973). *Acta Cryst.* A**29**, 308–309.

ROLLETT, J. S. (ed.) (1965). *Computing Methods in Crystallography*, p. 19. Oxford: Pergamon Press.

TIMOSHENKO, S. P. & YOUNG, D. H. (1965). *Theory of Structures*, 2nd ed. New York: McGraw-Hill.