Proc. 10th Int. Conf. on Mathem. and Computer Modelling and Scientific Computing, Boston, July 1995, published in: Mathematical Modelling and Scientific Computing, vol. 6, pp.325-330, 1996

SYSTEMATIC GENERATION OF MODELS FOR CRYSTAL STRUCTURES

Hans–Joachim Klein

Institut für Informatik und Praktische Mathematik, Universität Kiel, 24098 Kiel, Germany

ABSTRACT

A method is given for the systematic generation of graphs representing known or hypothetical polyhedral topologies of crystals. The method is complete for some interesting types of polyhedral topologies insofar as all topologies which are possible under reasonable assumptions concerning crystal properties are generated. Restriction modules can be coupled with the enumeration algorithm such that only graphs are obtained representing topologies with specific properties. For tetrahedral topologies, a technique is described for the embedding of graphs into three–dimensional Euclidean space.

KEYWORDS

crystallography, prediction of structures, polyhedral networks, polyhedral topologies, symmetries, periodic graphs, graph enumeration, three–dimensional embedding

INTRODUCTION

It is considered an interesting challenge to predict possible structures of inorganic crystalline solids from the knowledge of their chemical composition (Pannetier *et al.*, 1990) or from specified topological characteristics (O'Keeffe and Brese, 1992). Usual approaches to this problem start with initial coordinates of atoms and some parameters of a unit cell. They derive a potential arrangement of atoms by searching for minima with respect to a given energy model or by optimizing arrangements with respect to given rules using techniques like simulated annealing (Kirkpatrick *et al.*, 1983).

In this paper, we propose a different approach which is based on the modelling of atoms and their bonds in infinite perfect crystals by graphs. This modelling is usually done at the level of so–called coordination polyhedra. Coordination polyhedra are strongly bonded complexes of atoms with one atom located in the centre and adjacent atoms determining the corners of a polyhedron. Neighbouring polyhedra may be connected by sharing one or more atoms. Complex polyhedral networks can result from these connections.

For many crystals a single type of polyhedron and a single type of connection are sufficient for structure modelling, thus allowing a simple description of polyhedral networks by infinite graphs. The vertices of these graphs represent polyhedra, the edges represent connections between polyhedra. Because of the periodicity of crystals, each graph can be represented by a finite three–periodic

graph (Chung *et al.*, 1984, Goetzke *et al.*, 1988). Using operations of the so–called space group of a crystal as labels of edges it is possible to transform such a periodic graph into a digraph with vertices corresponding to points of a not necessarily unique fundamental domain called asymmetric unit.

This kind of minimal description is used as a basis for an enumeration algorithm generating models of possible crystal structures as follows: given an asymmetric unit of a space group and upper limits for the number and degree of vertices, graphs are produced for each admissible labeling of vertices and edges by so–called Wyckoff positions and symmetry operations, resp. Natural neighbourhood conditions for vertices involve restrictions on the choice of symmetry operations such that the number of labels to be considered is always finite.

Even for a small number of vertices and edges the number of different graphs is very large for most space groups. However, the method allows restrictions to be integrated such that it is possible to tailor the enumeration process to specific classes of crystals. Topological properties such as the occurrence and configuration of rings (Goetzke and Klein, 1991) may be checked during enumeration by efficient algorithms.

For graphs representing tetrahedral frameworks of silicates, we have developed an embedding algorithm which places vertices into the centre of gravity of all its neighbours and which corrects arrangements of atoms using knowledge of interatomic distances and bond angles.

REPRESENTATION OF POLYHEDRAL NETWORKS BY GRAPHS

The description of crystals is usually based upon the idea of a perfect structure. Such an ideal structure can be understood as an infinite three–dimensional space (crystal space) composed of cells which are arranged in a grid–like manner. Since all cells are considered to be identical in their atomic arrangement, a single specimen called unit cell is sufficient for the description of the atomic structure of a perfect crystal.

The cells of a crystal can be identified with elements of $\mathbb{Z}^3$ in the following way: the natural system spanned by the three linear independent vectors of the unit cell is considered as the reference system for the position data of all atoms. The coordinates of atoms in the unit cell are elements of $Z_0 = \{(y_1, y_2, y_3) \in \mathbb{R}^3 |\ 0 \leq y_1, y_2, y_3 < 1\}$. Let $(y_1, y_2, y_3) \in \mathbb{R}^3$ be the coordinates of an arbitrary atom A of a perfect crystal. Then its corresponding cell is given by the function $z|\ \mathbb{R}^3 \to \mathbb{Z}^3,\ z(y_1, y_2, y_3) = (\lfloor y_1 \rfloor, \lfloor y_2 \rfloor, \lfloor y_3 \rfloor)$.
An atom A' is called translationally equivalent to A if the difference of the coordinates of A and A' is an element of $\mathbb{Z}^3$. The coordinates of the atom $u(A)$ in the unit cell which is translationally equivalent to A are given by $(y_1 - \lfloor y_1 \rfloor, y_2 - \lfloor y_2 \rfloor, y_3 - \lfloor y_3 \rfloor)$.

Chemical bonds in materials are often modelled with the help of coordination polyhedra. Each polyhedron represents the bonds between a central atom and immediate neighbouring atoms. Two adjacent polyhedra may be connected by having one or more peripheral atoms in common. Thus, they may share a corner, an edge, or a face. Complex polyhedral networks can result from these connections. If geometric information such as interatomic distances and bond angles is not of interest, polyhedral networks are sometimes called polyhedral topologies.

In this paper, we consider polyhedral networks (topologies) in crystals consisting of a single type of polyhedron and a single type of connection between polyhedra. For many classes of crystals this kind of polyhedral network is well suited for the description of structures. Our approach, however, can be extended to arbitrary networks. The restriction to homogeneous networks allows to represent the main features of each polyhedral network P by an infinite polyhedra graph G_P in the following way: the vertices of G_P represent the polyhedra of P; there is a function *pos* which

assigns to each vertex v the coordinates of the atom in the centre of the polyhedron corresponding to v; the edges of G_P are determined by the connections between polyhedra in P. What we loose by this form of representation are the coordinates of the peripheral atoms.

Polyhedral networks of crystals may consist of a finite or an infinite number of connected parts. The connected parts may be finite or infinite, too, forming multiple polyhedra, rings, chains, layers, or frameworks. Thus, a polyhedra graph can have a finite or an infinite number of finite or infinite connected units. However, because of translational equivalence it is possible to represent each polyhedra graph by a finite digraph.

Definition : A direction–labeled graph $G_{dl} = (V, E, pos, \alpha, \omega, \delta)$ is a finite, labeled digraph with set V of vertices, set E of edges, and functions $pos : V \to Z_0$, $\alpha, \omega : E \to V$, and $\delta : E \to \mathbb{Z}^3$ such that the following holds: an edge $e \in E$ is directed from the start–vertex $\alpha(e)$ to the end–vertex $\omega(e)$ and labeled by $\delta(e)$; there is exactly one complementary edge $\tilde{e} \in E$ different from e with $\alpha(\tilde{e}) = \omega(e)$, $\omega(\tilde{e}) = \alpha(e)$ and $\delta(\tilde{e}) = -\delta(e)$.

The direction–labeled graph (dl–graph) of a polyhedral network P of a crystal S is derived from the corresponding polyhedra graph G_P with set V_P of vertices and set E_P of edges as follows:

$V := \{v \in V_P |\ pos(v) \in Z_0\}$

$E := \{e |\ e \textit{ uniquely determined by } (\exists \{v, v'\} \in E_P)(\alpha(e) = v \wedge \omega(e) = u(v') \wedge \delta(e) = z(pos(v')))\}$

The vertices of a dl–graph G_{dl} correspond to the polyhedra with central atom in the unit cell of a crystal S. For each vertex v_i of G_{dl} there is an edge directed to a vertex v_j if the corresponding polyhedra p_i, p_j are connected or if p_i is connected to a polyhedron $\tilde{p}_j$ with central atom in a neighbouring cell of the unit cell and if (the central atom of) $\tilde{p}_j$ is translationally equivalent to (the central atom of) p_j. The edge with start–vertex v_i and end–vertex v_j is labeled by $(0,0,0)$ in the first case and by an element of $\{0, 1, -1\}^3$ identifying the cell of $\tilde{p}_j$ in the second case. There are redundant edges in G_{dl} because each edge has a complementary edge. This redundancy is useful for the formulation of algorithms.

The following figure shows part of a tetrahedral layer in semenovite and the corresponding connected unit of its dl–graph (there is another layer in semenovite which is symmetrically equivalent to the one shown in the figure, i.e. the dl–graph of the tetrahedral network of semenovite has two connected units which are identical except for the function *pos*). In the drawing of the graph, a pair of complementary edges $\bullet \xrightarrow{t} \bullet$, $\bullet \xleftarrow{-t} \bullet$, $t \in \mathbb{Z}^3 \setminus \{(0,0,0)\}$, is represented by a single edge $\bullet \overset{-t+}{\text{---}} \bullet$. The edges without labels represent pairs of complementary edges labeled by $(0,0,0)$.

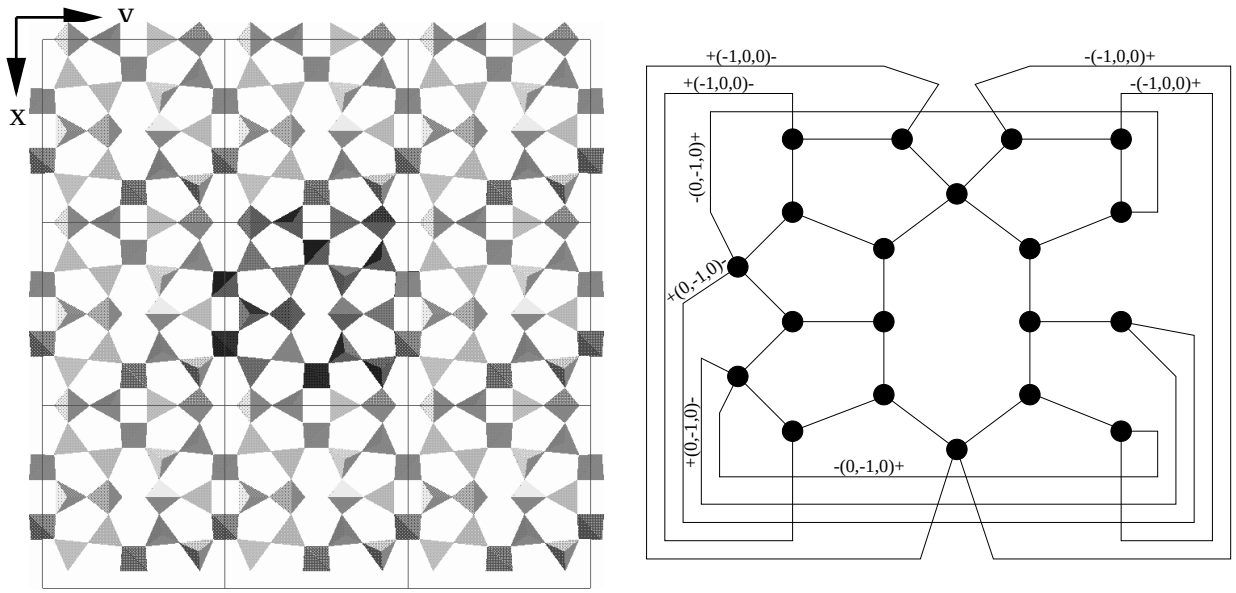

Fig. 1: Part of the tetrahedral layer found in semenovite and its representation by a dl–graph

Direction–labeled graphs are well suited for the formal definition and computation of those properties of polyhedral networks which refer directly to translational symmetry. The dimensionality of a network, for example, is the number of elements in a basis of the cycle space of its dl–graph. These graphs, however, are in general not the most comprehensible form to represent polyhedral networks. The reason for this are symmetries other than translational symmetries which can be found in most crystals. As an example, consider the crystal semenovite. In this structure mirror planes, a centre of inversion, twofold rotation and screw axes, and diagonal glide reflection planes can be found. Some of these symmetries can be perceived in the part of the layer shown in Fig. 1.

The symmetry operations of a three–dimensional crystal pattern form a group called the space group of the crystal. Because of the infinity of crystal spaces and because of translational symmetries, this group is always infinite. For each space group there is a not necessarily unique fundamental domain called asymmetric unit. Each point in a crystal space can be obtained by applying symmetry operations of its space group to a point in such an asymmetric unit. An asymmetric unit is minimal insofar as none of its points can be removed without loosing the generating property. It is not necessarily part of a single unit cell.
A three–dimensional crystal space can be considered as being composed of 'copies' of an arbitrary asymmetric unit instead of 'copies' of a unit cell. Each such copy can be obtained by applying a symmetry operation of the space group to all points of the asymmetric unit. Conversely, symmetry operations of the space group can be used to 'fold' a dl–graph in such a way that only those polyhedra are represented as vertices whose central atom is located in the asymmetric unit. Edges have to be labeled appropriately.

Definition: Let G_{dl} be the dl–graph of a polyhedral network P of some crystal S and let A_s be an arbitrary but fixed asymmetric unit of S. The symmetry–labeled graph G_{sl} of P with respect to A_s (sl–graph of P) is a finite, labeled digraph with a set V_{as} of vertices, a set E_{as} of edges labeled by symmetry operations, and a function $pos : V_{as} \to \mathbb{R}^3$. It is derived from G_{dl} as follows:
V_{as} contains for each vertex v of G_{dl} exactly one vertex v_{as} with: $pos(v_{as}) \in A_s$ and $pos(v)$ can be obtained from $pos(v_{as})$ by application of a symmetry operation of the space group of S. V_{as} contains no other vertices. Two vertices v_1, $v_2 \in V_{as}$ are connected by an edge with start–vertex v_1, end–vertex v_2, and symmetry operation σ as label iff there is an edge e in G_{dl} such that the following holds: $\alpha(e) = \tilde{v}_1$, $\omega(e) = \tilde{v}_2$, and symmetry operations $\tilde{\sigma}_1$, $\tilde{\sigma}_2$ exist in the space group of S with $pos(\tilde{v}_1) = \tilde{\sigma}_1(pos(v_1))$, $pos(\tilde{v}_2) = \tilde{\sigma}_2(pos(v_2))$, and $\sigma = \tilde{\sigma}_1^{-1} \circ \delta(e) \circ \tilde{\sigma}_2$.

The next figure shows the sl–graph derived from the dl–graph in Fig. 1. The choice of the vertices for the asymmetric unit can be inferred from the drawings. For the labeling of edges the following symmetry operations are used: reflection $(\sigma_m, \sigma_{m'})$, screw rotation (σ_{2_1}), and a translation by $(0, -1, 0)$ (σ_t). Edges of the graph labeled with the identity are drawn without label.

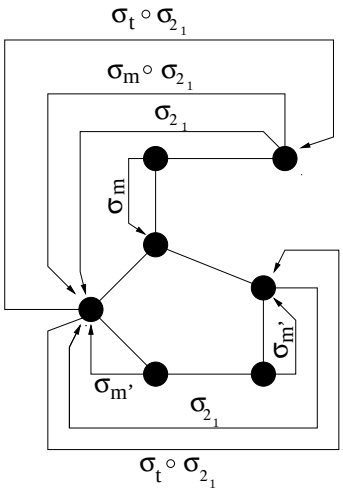

Fig. 2: Symmetry–labeled graph for the tetrahedral layers of semenovite

As we have demonstrated in the preceding section, each polyhedra graph can be represented by an sl–graph. Since there is no redundancy in this representation, sl–graphs are well suited as a basis for enumerating polyhedra graphs. In an enumeration process it makes no sense to assign real coordinates to vertices for obvious reasons. Therefore the function *pos* has to be changed. It cannot be dropped completely since symmetry operations have to be applied to coordinates of vertices when an sl–graph is unfolded to a dl–graph. For the application of symmetry operations it is sufficient to have general coordinates (x, y, z). However, it is more favourable to use information about the effect of the application of symmetry operations on points in the asymmetric unit. Each point belongs to a so–called Wyckoff position, which is a set of points having conjugate site symmetry groups. The site symmetry group of a point is the set of symmetry operations which leave a point invariant. The number of different Wyckoff positions is finite for each space group (≤ 27). Because of the equivalence of its elements with respect to the symmetry operations of the underlying space group, it is sufficient to choose a single representative for each Wyckoff position as a possible value of the function *pos*. In Hahn (1992), the Wyckoff positions of all space–group types are documented. A letter is assigned to each Wyckoff position being unique for each type. We use this letter for referencing Wyckoff positions. Let us call the type of graph resulting from these modifications potential sl–graph.

In order to reduce the number of potential sl–graphs which cannot contribute to the set of hypothetical polyhedral networks, some restrictions can be integrated into the enumeration algorithm for potential sl–graphs. First, symmetry operations used as labels of edges in a potential sl–graph have to be elements of the same space–group type. This is a direct consequence of the uniqueness of space groups for crystal structures. The second restriction concerns the way symmetry operations can be combined as labels in a graph. There are interdependencies and incompatibilities of labels implied by the group structure of symmetry operations. The third restriction has a less formal but none the less safe background. It is a consequence of the nature of bonds: interatomic bonds are assumed to be possible only in certain ranges of distances; therefore it makes no sense to consider symmetry operations as labels which represent bonds between polyhedra having too large distances. As a consequence of this natural condition, only a finite subset of symmetry operations of each (infinite) space–group type is relevant for the enumeration of potential sl–graphs. In order to determine this subset, we use so–called neighbourhoods. The first neighbourhood of an asymmetric unit A_s is the set of 'copies' of A_s sharing one or more points with A_s. The shape of asymmetric units can be rather complicated for space groups with higher symmetry. Furthermore, in space groups with lower symmetry where asymmetric units are parallelepipedes, the spanning vectors can be quite different in length. Therefore we allow to parameterize the enumeration process and to consider not only symmetry operations generating positions in the first neighbourhood but also operations leading into the second or higher neighbourhoods in specific directions. Because of the restrictions concerning the length of interatomic bonds, there are, however, only few space–group types for which higher neighbourhoods have to be taken into account.

The set of potential sl–graphs contains all representatives of homogeneous polyhedral topologies. If the number of polyhedra which are not translational equivalent is restricted, then for each type of polyhedron the set of potential sl–graphs to be considered as representatives of possible topologies is finite and can be computed effectively.

In order to make the enumeration goal–directed, i.e. to tailor it to graphs representing polyhedral topologies with specific properties, the enumeration algorithm can be parameterized. The following parameters are possible as input: • the space group type, • the number of neighbourhoods in specific directions, • the number of vertices (polyhedra) in the asymmetric unit (min, max) and in the unit cell (max), • the degree of vertices (min, max), • the dimensionality of the network, and • Wyckoff positions of vertices.
Additionally, there is an interface to combine the algorithm with restriction modules checking

properties such as the existence of rings of a certain type.
Some of the parameters reduce the number of sl–graphs to be generated. The maximal number of vertices in the unit cell, for example, can restrict the choice of Wyckoff positions for the vertices. Other parameters, dimensionality for example, can be tested after the unfolding of a given sl–graph to a dl–graph has been performed. An sl–graph G_{sl} can be unfolded to a dl–graph by applying for each edge e of G_{sl} the symmetry operations of the space–group type of G_{sl} to $pos(\alpha(e))$ and $\delta(e)(pos(\omega(e)))$. At the level of dl–graphs, some interesting properties can be tested not referring to coordinates of atoms, i.e. being of topological nature. Among these properties is the ring configuration of polyhedral topologies (see Goetzke and Klein, 1991) which plays an important role for the analysis of several classes of crystals.

EMBEDDING OF TETRAHEDRAL TOPOLOGIES

Direction–labeled graphs generated by the enumeration algorithm contain topological information and the Wyckoff position of each vertex. Moreover, the space group type is known for which the graph has been produced. Depending on this type, there is some information on angles between axes and the ratio of the lengths of axes. For the embedding of a polyhedral graph into three–dimensional Euclidean space we have to determine cell data, i.e. the length of each axis and angles between axes for those space group types for which these angles are not fixed. The position data of vertices have to be determined with respect to this vector space.

For tetrahedral topologies, which are very frequent, we have developed an embedding algorithm proceeding as follows: first, each vertex is embedded into the centre of gravity of its neighbours. Because of translational symmetry, this can be done by solving a simple system of linear equations. Sometimes, collisions may occur which have to be treated appropriately. In the second phase of the algorithm, positions of atoms and cell parameters are optimized with respect to interatomic distances and bond angles. The concrete values depend on the kind of atoms assumed. Optimization can be done alternatively by using simulated annealing or gradient methods. Tests with a large number of silicates have shown this proceeding to be very promising.

REFERENCES

Chung S.J., T. Hahn, and W.E. Klee (1984). Nomenclature and Generation of Three–Periodic Nets: the Vector Method. *Acta Cryst.*, A 40, 42-50.

Goetzke, K., H.–J. Klein, and P. Kandzia (1988). Automatic Crystal Chemical Classification of Silicates Using Direction–labeled Graphs. In: *Proc. Int. Workshop on Graph–Theoretic Concepts in Computer Science* (H. Göttler and J. Schneider, eds.), Lecture Notes in Computer Science, 314, 242-254, Springer, Berlin.

Goetzke, K. and H.–J. Klein (1991). Properties and Efficient Algorithmic Determination of Different Classes of Rings in Finite and in Infinite Polyhedral Networks. *J. of Non–Crystalline Solids*, 127, 215-220.

Hahn T. (ed.) (1992). *International Tables for Crystallography, Vol. A, Space–Group Symmetry.* Kluwer Academic Publishers, Dordrecht.

Kirkpatrick S., C.D. Gelatt, and M.P. Vecchi (1983). Simulated Annealing. *Science*, 220, 671-680.

O'Keeffe M. and N.E. Brese (1992). Uninodal 4–Connected 3D Nets. I. Nets without 3– or 4–Rings. *Acta Cryst.*, A48, 663-669.

Pannetier J. et al. (1990). Prediction of Crystal Structures from Crystal Chemistry by Simulated Annealing. *Nature*, 346, 343-345.