**Classification and use of topological data**

By V. A. Blatov, D. M. Proserpio and R. M. Hanson

**1. Introduction**

The crystallographic data obtained from experiment or theoretical modeling contain no explicit information on the crystal structure connectivity. However, this information, which might describe all sorts of interatomic bonding as well as the methods of connection of complex structural units, is crucial for many tasks of crystal chemistry and materials science, such as classification, comparative analysis, modeling of crystal growth and assembling, as well as the search for correlations among crystal composition, structure and properties. Similar to molecular graphs, which are widespread in organic chemistry, crystal structure can be represented as a three-periodic graph, which is usually called a *net* (Delgado-Friedrichs & O’Keeffe, 2005; Delgado-Friedrichs *et al*., 2005). Unlike molecular graphs, nets are infinite and require special methods of description. Such descriptions have become increasingly common in routine crystal structure determinations and usually follow the conventional crystallographic description, which includes atomic coordinates, space group, unit cell dimensions, and other geometrical information, but does not concern the structure connectivity. Many topological descriptors have been proposed for nets. All of these descriptors are determined from the initial crystallographic data using algorithms. Software implementing these algorithms, such as *Gavrog Systre* (Delgado-Friedrichs & O’Keeffe, 2003) or *ToposPro* (Blatov *et al*., 2014), has made topological analysis available for the general crystallographic community and has already led to the development of special databases (O'Keeffe *et al*., 2008; Blatov *et al*., 2014) that collect the topological parameters of crystal structures. This work will be facilitated by an extension of the CIF format that makes it possible to store topological information along with a crystal structure and to describe the crystal architecture at different levels of organization.

Although the core CIF dictionary contains some topological information in the CHEMICAL\_CONN, GEOM\_BOND, GEOM\_HBOND, and GEOM\_CONTACT categories, this information is limited and does not cover all topological aspects of a crystal. For example, while the core CIF dictionary does have the ability to describe simple covalent and hydrogen bonds as well as van der Waals interactions, it cannot be used to describe more nuanced bonding, such as halogen-halogen bonds or other special bond types. Moreover, structures that contain polyatomic *building units*, such as molecules, metal complex groups, or clusters, can be simplified to form “nodes” and “links” of an underlying net, which bears the general topological motif of these units, and have positions that may not be atom positions (Alexandrov *et al*. 2011; Bonneau *et al.* 2018; Barthel *et al.* 2018; Shevchenko & Blatov, 2021). The topoCIF dictionary introduces a standard CIF-based description of all of these varied interactions and net topology.

**2. Dictionary design considerations**

**2.1 Overall format** The topoCIF dictionary can be expressed in either CIF 1 or CIF 2 format. Where there is a difference, it is in the use of CIF 2 lists for just a few data items. The discussion below lists the CIF 2 data keys; ⱡ indicates the availability of CIF 1 equivalents, which are discussed where appropriate.

**2.2 Categories and scope**

The topoCIF dictionary contains five new categories for the description of connectivity and topological properties of periodic nets, namely TOPOL\_NET, TOPOL\_NODE, TOPOL\_LINK, and TOPOL\_ATOM, and TOPOL\_TILING. These five categories can be divided into two groups: (i) descriptors of crystal structure connectivity (TOPOL\_NODE, TOPOL\_LINK, and TOPOL\_ATOM), and (ii) descriptors of topological properties of a periodic net (TOPOL\_NET and TOPOL\_TILING). The descriptors of the first group can be computed directly from the initial crystallographic data; the descriptors of the second group are determined based on that connectivity information. Thus, the topoCIF dictionary enables one to generate two new levels of the crystal structure description: a connected net without any further detailing and a topological description of that connected net. The description of tiling is minimal in this first release of the dictionary. The dictionary can be expanded in a future version to provide a more complete description of tiles and their faces and edges. A sixth category, TOPOL\_ENTANGL, is proposed but has been left empty until the terminology on entanglements gains consistency, but since the corresponding terminology is still under development, only conventional terms and descriptors defined by Delgado-Friedrichs & O’Keeffe (2005) and Blatov *et al*. (2010) are included in this first version of the dictionary.

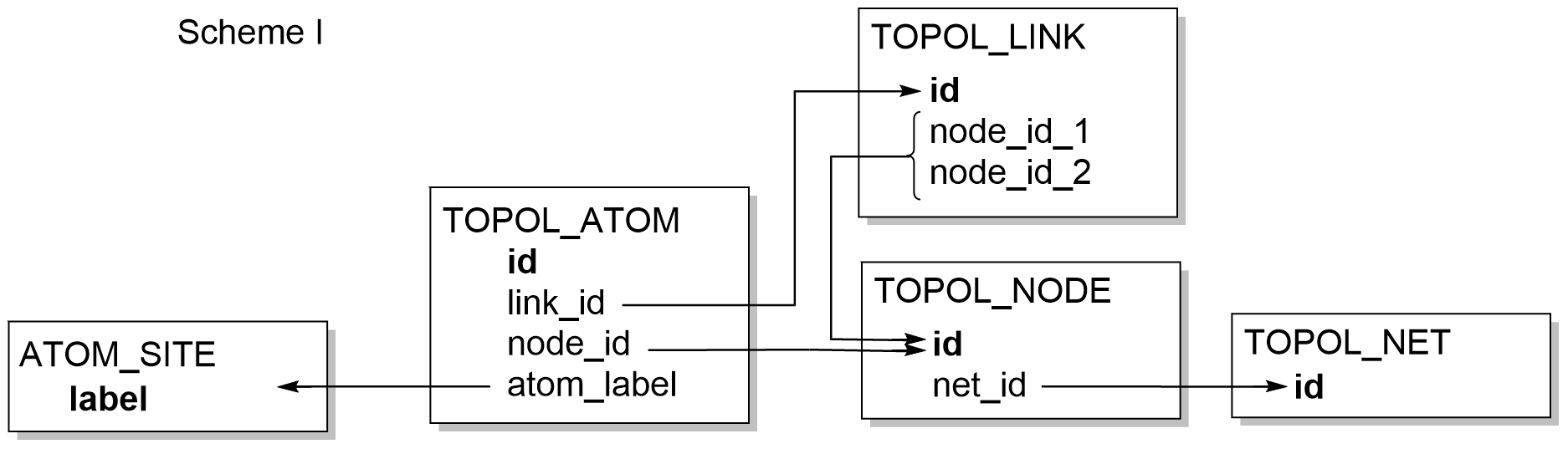
**3. Classification of data definitions**

**3.1. Underlying net description**

The *underlying net* is the net of centroids of structural units. The edges of the net represent the links between these building units. The underlying net describes the topology (Francl, 2009) of a particular representation of a crystal structure (Eon, 2016). This representation is said to be *complete* if the underlying net coincides with the initial structure, that is, if the building units coincide with the atoms described in ATOM\_SITE , and *partial* if not. For example, diamond and SiO2 (cristobalite) can be represented by the same underlying net **dia**. For diamond, this net is complete, because all atoms and bonds we would normally associate with the diamond structure are represented by the net. But for cristobalite, **dia** is only a partial net, since it only represents the silicon atoms. The oxygen atoms can be considered to be bridges (links) between silicon atoms, but those links are represented just by edges of the **dia** net.

All topological data in the TOPOL category describe one or more (complete or partial) topological representations of the crystal structure given in the form of a periodic net. The nodes of this net either coincide with the atoms or correspond to the centers of building units. The two main categories, TOPOL\_LINK and TOPOL\_NODE, describe the underlying net and its connectivity without any specific reference to atoms. The relationship of links and nodes to atoms listed in ATOM\_SITE are established in the TOPOL\_ATOM category. The TOPOL\_NET category is necessary either if the crystal structure contains several non-equivalent nets, or if one needs to describe several representations of the same crystal structure. This category can also be used for the description of overall net features.

Scheme I shows the overall relationship among TOPOL\_NET, TOPOL\_LINK, TOPOL\_NODE, and TOPOL\_ATOM. TOPOL\_ATOM provides references to ATOM\_SITE atoms associated with nodes and links. Note that only TOPOL\_ATOM references ATOM\_SITE directly.



A topoCIF file can describe any number of nets. The TOPOL\_NET category (see part 3.2) describes each individual net, and \_topol\_net.id is referenced from TOPOL\_NODE. Note that since links refer to nodes, and TOPOL\_ATOM items refer to nodes as well as links, any atom that is used in more than one net must be listed twice in TOPOL\_ATOM, in each case referring to a different link or node. Note also that it is possible, depending upon the symmetry, for a given “net” indexed in TOPOL\_NET to consist of two or more *interpenetrating* but not *interconnected* nets. (Examples 1, 2, and 4 illustrate such nets.) The presence of such interpenetration is identified by the data item \_topol\_net.z\_number, which indicates the number of interpenetrating nets. However, nothing in this version of the dictionary (other than the connectivity itself) distinguishes among such nets.

In the discussion below, bold face indicates the category key, a unique integer, and arrow (→) indicates a reference to another category’s key having the same value.

**TOPOL\_NODE**

The TOPOL\_NODE category describes the minimum number of nodes associated with the asymmetric unit that need to be defined in order for the application of symmetry operators to generate the full periodic net. The category involves the following data items:

\_topol\_node.coordination\_sequence

\_topol\_node.coordination\_sequence\_plain

\_topol\_node.extended\_point\_symbol

\_topol\_node.fract\_x

\_topol\_node.fract\_y

\_topol\_node.fract\_z

**\_topol\_node.id**

\_topol\_node.label

\_topol\_node.net\_id → \_topol\_net.id

\_topol\_node.point\_symbol

\_topol\_node.symmetry\_multiplicity

\_topol\_node.vertex\_symbol

\_topol\_node.Wyckoff\_symbol

The \_topol\_node.id data item is the only a required tag, as it is the category key, referenced by TOPOL\_LINK and TOPOL\_ATOM. Its integer value must be unique and typically the series starts with the value 1. In fact, in many cases, \_topol\_node.id is the only TOPOL\_NODE data item necessary (see Examples 1, 4, and 6). The data items \_topol\_node.fract\_x, \_topol\_node.fract\_y, and \_topol\_node.fract\_z are symmetry-transformed fractional coordinates. They are needed only when the node refers to a non-atomic position that is not characterizable as a centroid, such as the center of a channel. If a node represents more than one atom from ATOM\_SITE, fractional coordinates given in TOPOL\_NODE give the *author-preferred* locus of the node for rendering purposes only. Otherwise, the centroid of those atoms is generally to be assumed. \_topol\_link.net\_id is required only if the TOPOL\_NET category is present and there is more than one net in the structure or in the description of the structure representation. Note that any number of nodes may correspond with a given net, but each node must correlate with only one net. For nodes associated with single atoms of ATOM\_SITE, \_topol\_node.symmetry\_multiplicity and \_topol\_node.Wyckoff\_symbol should not be present, as they would simply duplicate information in the ATOM\_SITE category.

**TOPOL\_LINK**

The TOPOL\_LINK category defines all of the connectivity of the nodes and allows for the application of additional symmetry operations on the nodes defined in TOPOL\_NODE and TOPOL\_ATOM. The category includes the following data items:

\_topol\_link.distance

**\_topol\_link.id**

\_topol\_link.label

\_topol\_link.multiplicity

\_topol\_link.node\_id\_1 → \_topol\_node.id

\_topol\_link.node\_id\_2 → \_topol\_node.id

\_topol\_link.order

\_topol\_link.special\_details

\_topol\_link.symop\_id\_1 → \_space\_group\_symop.id

\_topol\_link.translation\_1ⱡ

\_topol\_link.symop\_id\_2 → \_space\_group\_symop.id

\_topol\_link.translation\_2ⱡ

\_topol\_link.type

\_topol\_link.Voronoi\_solid\_angle

The \_topol\_link .id item is required, as it is the unique category key. It is an integer, typically starting with 1. Similar to the GEOM\_BOND category, TOPOL\_LINK allows for a general reference to an ATOM\_SITE atom along with a symmetry operation and a translation to be applied to that atom’s reference position. The syntax, however, is slightly different. Like \_geom\_bond.symmetry\_\*, \_topol\_link.symop\_id\_\* references \_space\_group\_symop.id, but the added translational component is indicated as a list [ i j k ] with [1 0 0] representing a unit translation along the *a* axis. (This allows for a more general representation than the “555” notation used in \_space\_group\_symop.id.) Defaults for \_topol\_link.symop\_id\_\* and \_topol\_link.translation\_\* are 1 and [0 0 0], respectively. The CIF 1 equivalents for \_topol\_link.translation\_\* are \_topol\_link.translation\_\*\_x, \_topol\_link.translation\_\*\_y, and \_topol\_link.translation\_\*\_z, where \* is 1 or 2. The data items \_topol\_link.distance, \_topol\_link.multiplicity, \_topol\_link.order, \_topol\_link.special\_details, \_topol\_link.type, and \_topol\_link.Voronoi\_solid\_angle are additional optional data items that describe characteristics of the link. For example, \_topol\_link.Voronoi\_solid\_angle , designates the solid angle of an interatomic contact A-X, which corresponds to a common face of the Voronoi polyhedra of A and X atoms (Blatov, 2004; Fig. 1). These items are described more fully in the dictionary descriptions.

A picture containing athletic game, sport

Description automatically generated

**Figure 1:** The solid angle of the shaded Voronoi polyhedron face in the body-centered cubic lattice is equal to the shaded segment of the unit sphere being cut off by the pyramid with the Voronoi polyhedron atom A at the vertex and the face in the base. The shaded face is shared between Voronoi polyhedra of A and X atoms and corresponds to the bond A-X.

**TOPOL\_ATOM**

The TOPOL\_ATOM category correlates atoms from ATOM\_SITE to nodes and links. This correlation allows the definition of a node that corresponds with a group of atoms of the crystal structure, or the definition of a link that itself (irrespective of its end points) represents a group of atoms (*building units*, Alexandrov *et al*. 2011; Bonneau *et al.* 2018; Shevchenko & Blatov, 2021) Thus, TOPOL\_ATOM describes the chemical constitution of nodes and links, and TOPOL\_ATOM, TOPOL\_NODE, and TOPOL\_LINK fully describe the connectivity, geometric, and chemical aspects of a net. The TOPOL\_ATOM category is quite simple, just:

\_topol\_atom.atom\_label

\_topol\_atom.element\_symbol

**\_topol\_atom.id**

\_topol\_atom.link\_id → \_topol\_link.id

\_topol\_atom.node\_id → \_topol\_node.id

\_topol\_atom.symop\_id

\_topol\_atom.translationⱡ

The \_topol\_atom .id data item is required (even if it is not referenced), as it is the unique category key and has no default value. It is an integer, typically starting with 1. \_topol\_atom.atom\_label is case-sensitive and required, as it is the reference to ATOM\_SITE. One or both of \_topol\_atom.node\_id and \_topol\_atom.link\_id is required, pointing to the node and/or link associated with this atom. Note that this allows any number of \_topol\_atom data items to be associated with one and only one node and one and only one link (possibly both). The data item \_topol\_atom.element\_symbol is strongly recommended, as there is no other location in a CIF file where a specific element symbol is required. If for some reason the ATOM\_SITE atom does not relate to a specific element (due to occupational disorder or some other reason), this field should be left as unquoted ‘.’.

**3.2. Underlying net topological properties**

The topological properties of underlying nets described by the TOPOL\_NODE, TOPOL\_LINK and TOPOL\_ATOM categories are collected in the remaining TOPOL categories, namely TOPOL\_NET, TOPOL\_TILING, and TOPOL\_ENTANGL. These categories are discussed below.

**TOPOL\_NET**

The TOPOL\_NET category enables one to describe a crystal structure in different representations, which depend on the selected structural units and levels of interatomic interaction (Alexandrov *et al*., 2011). This approach can resolve the problem of ambiguity in the description of the crystal structure connectivity: the crystal structure can be treated in different ways and described by several underlying nets listed in a TOPOL\_NET loop. If there is only one net, and that net has no data items of note, the TOPOL\_NET category may be omitted, provided it is not referenced in TOPOL\_NODE. However, if a net has defined characteristics, or if it is referenced, or if several nets are described, the TOPOL\_NET category is necessary. Its items include:

\_topol\_net.genus

**\_topol\_net.id**

\_topol\_net.label

\_topol\_net.overall\_topology\_EPINET

\_topol\_net.overall\_topology\_IZA

\_topol\_net.overall\_topology\_RCSR

\_topol\_net.overall\_topology\_SP

\_topol\_net.overall\_topology\_TOPOS

\_topol\_net.period

\_topol\_net.special\_details

\_topol\_net.td10

\_topol\_net.total\_point\_symbol

\_topol\_net.z\_number

The \_topol\_net.id, a serial integer, is required if there is more than one net or the net is to be referenced in TOPOL\_NODE, as it is the category key. \_topol\_net.label allows for an optional arbitrary name for the net. The overall topology of the net can be designated in accordance with five mostly recommended nomenclatures (RCSR, EPINET, SP, TOPOS or IZA), which have special tags:

* RCSR lower-case bold three-letter symbols, see O'Keeffe *et al*. (2008) and http://rcsr.net/ for details. Example: **dia** (see http://rcsr.net/nets/dia)
* EPINET *sqc*XXXXX symbols, see Ramsden *et al*. (2009) and http://epinet.anu.edu.au/ for details.
* SP (sphere packing) symbols *k*/*m*/*fn* as proposed by Koch *et al*. (2006).
* TOPOS symbols N*D*n (Aman *et al*., 2014), where N is a sequence of degrees (coordination numbers) of all independent nodes; *D* is one of the letters *C*, *L*, or *T* designating the dimensionality of the net (*C* – chain, *L* – layer, *T* – three-periodic); n enumerates non-isomorphic nets with a given N*D* sequence. For instance, the symbol 3,3,4*T*3 denotes the 3rd (by the order) three-periodic trinodal net with two 3-coordinated and one 4-coordinated independent nodes. For finite (molecular) graphs the symbols N*M*k-n are used, where k is the number of vertices (atoms) in the graph.
* IZA-SC capital three-letter symbols for zeolites, see http://www.iza-structure.org/databases/ for details.

If the topology is designated in another nomenclature it should be described in \_topol\_net.special\_details, where an arbitrarily detailed description of the net can be provided.

Additionally, this category contains conventional topological descriptors of a net, such as net periodicity (\_topol\_net.period), topological density TD10 (\_topol\_net.td10), point symbol (\_topol\_net.total\_point\_symbol) and number of unconnected nets (*Z*) associated with this entry (\_topol\_net.z\_number). For more details and definitions see Delgado-Friedrichs & O’Keeffe (2005), Delgado-Friedrichs *et al*. (2005) and Blatov *et al*. (2010).

**TOPOL\_TILING**

This category describes a tiling, which is carried by the underlying net. In general, each underlying net can have its own tiling described in TOPOL\_TILING. As a rule, natural tiling (Blatov *et al*., 2007) is described, since it contains minimal cages, which can be combined to provide all other larger cages, including infinite channels. Items in TOPOL\_TILING include:

\_topol\_tiling.d\_size

\_topol\_tiling.dual

\_topol\_tiling.edges

\_topol\_tiling.faces

\_topol\_tiling.id

\_topel\_tiling.net\_id → \_topol\_net.id

\_topol\_tiling.signature

\_topol\_tiling.tiles

\_topol\_tiling.vertices

**TOPOL\_ENTANGL**

This category contains no items, because the descriptors of entanglements in crystal structures have not been standardized yet. However, the importance of entanglements in the topological description is acknowledged, and this category is expected to be developed in a future version of the dictionary.

**3.4 Examples**

In these seven examples, an ellipsis (…) is used to abbreviate the list of symmetry operations in the SPACE\_GROUP\_SYMOP category. All examples are available as CIF files at <https://github.com/COMCIFS/TopoCif/tree/master/examples>.

Example 1 describes the diamond network. The net requires just one node and one link, which, upon applying symmetry, generates the full periodic network.

Example 2 involves two nets, one of which includes a polyatomic node – the carbon and oxygen atoms of a carbonyl ligand – with a position that is given explicitly.

Example 3 involves a polyatomic node for which the position of the node is implied to be the centroid of the associated atoms.

Example 4 illustrates a \_topol\_net that consists of two interpenetrating nets with a singleatom linker (Cu).

Example 5, for the metal-organic framework MOF5, involves three distinct nets, requiring eleven nodes, ten links, and fifty atoms.

Example 6 illustrates the use of TOPOL\_\* in place of GEOM\_BOND for molecular structure, including hydrogen bonding.

Example 7 includes a description of tiling.

|  |
| --- |
| *Example 1. Connectivity of the diamond crystal structure. All nodes correspond to atoms in ATOM\_SITE. All links correspond to covalent bonds, so the underlying net corresponds to the atomic network.*  #\#CIF\_2.0  data\_example\_1  \_chemical\_formula.sum C1  \_cell.length\_a 3.567  \_cell.length\_b 3.567  \_cell.length\_c 3.567  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 45.38468  \_cell.formula\_units\_Z 8  \_space\_group.name\_H-M\_ref 'F d -3 m'  \_space\_group.name\_Hall '-F 4vw 2vw 3'  \_space\_group.IT\_number 227  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  192 3/4-z,1/2+y,1/4-x  \_atom\_site.label C1  \_atom\_site.fract\_x 0.12500  \_atom\_site.fract\_y 0.12500  \_atom\_site.fract\_z 0.12500  \_topol\_net.id 1  \_topol\_net.overall\_topology\_RCSR dia  \_topol\_node.id 1  \_topol\_node.net\_id 1  \_topol\_link.id 1  \_topol\_link.node\_id\_1 1  \_topol\_link.node\_id\_2 1  \_topol\_link.symop\_id\_1 1  \_topol\_link.translation\_1 [0 0 0]  \_topol\_link.symop\_id\_2 13  \_topol\_link.translation\_2 [0 0 0]  \_topol\_link.distance 1.5446  \_topol\_link.Voronoi\_solid\_angle 22.04  \_topol\_link.type v  \_topol\_link.order 1  \_topol\_link.multiplicity 16  \_topol\_atom.id 1  \_topol\_atom.node\_id 1  \_topol\_atom.atom\_label C1  \_topol\_atom.element\_symbol C  #End of data\_example\_1 |

|  |
| --- |
| *Example 2. Connectivity of atomic and underlying nets for an interpenetrating array of two LiCo(CO)4 networks. Both the atomic and the underlying nets are described in the TOPOL\_NET section. The atomic net consists of Li, C, O, and Co atoms, while the underlying net is built from three kinds of nodes: Li and Co atoms and carbonyl (CO) ligand, which are labeled ZA1, ZC1, and ZB1, respectively. Two possible options for describing the nodes are shown: the coordinates of ZA1 are specified by a reference (from TOPOL\_ATOM) to the Li1 atom, while the coordinates of ZC1 are specified explicitly, because a point is being chosen that is not the centroid of the C and O positions.*  #\#CIF\_2.0  data\_example\_2  \_database.code\_CSD FUBYUW  \_chemical.name\_systematic 'lithium tetracarbonyl-cobalt'  \_chemical\_formula.sum 'C4 Co1 Li1 O4'  \_cell.length\_a 5.542  \_cell.length\_b 5.542  \_cell.length\_c 5.542  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 170.2157  \_cell.formula\_units\_Z 1  \_space\_group.name\_H-M\_alt 'P -4 3 m'  \_space\_group.name\_Hall 'P -4 2 3'  \_space\_group.IT\_number 215  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  24 -z,y,-x  loop\_  \_atom\_site.label  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  Li1 0.00000 0.00000 0.00000  C1 0.31850 0.31850 0.31850  O1 0.19920 0.19920 0.19920  Co1 0.50000 0.50000 0.50000  loop\_  \_topol\_net.id  \_topol\_net.label  \_topol\_net.z\_number  \_topol\_net.special\_details  \_topol\_net.overall\_topology\_TOPOS  1 Net\_1 2 'Atomic network' 'Unknown'  2 Net\_2 2 'Underlying net with carbonyl ligands as nodes' '2,4T3'  loop\_  \_topol\_node.id  \_topol\_node.label  \_topol\_node.net\_id  \_topol\_node.fract\_x  \_topol\_node.fract\_y  \_topol\_node.fract\_z  1 Li1 1 . . . # Li  2 C1 1 . . . # C  3 O1 1 . . . # O  4 Co1 1 . . . # Co  5 ZA1 2 . . . # Li  6 ZB1 2 0.25036 0.25036 0.25036 # (CO)  7 ZC1 2 . . . # Co  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.distance  \_topol\_link.type  1 1 3 1.9121 v # Li1-O1  2 2 3 1.1452 v # C1-O1  3 2 4 1.7422 v # C1-Co1  4 5 6 2.4032 gl # Li1-(CO)  5 6 7 2.3963 gl # (CO)-Co  loop\_  \_topol\_atom.id  \_topol\_atom.node\_id  \_topol\_atom.atom\_label  \_topol\_atom.element\_symbol  1 1 Li1 Li  2 2 C1 C  3 3 O1 O  4 4 Co1 Co  5 5 Li1 Li  6 6 C1 C  7 6 O1 O  8 7 Co1 Co  #End of data\_example\_2 |

|  |
| --- |
| *Example 3. The underlying net of the calcite (CaCO3) crystal structure. The nodes of the net correspond to carbonate groups (node ZA1) and Ca atoms (node ZA2). Positions of the nodes are not specified, as they can be determined directly from the centroids of the associated atoms. The net has the NaCl topology (***pcu-b** *in the RCSR nomenclature).*  #\#CIF\_2.0  data\_example\_3  \_database.code\_ICSD 16710  \_chemical.name\_systematic 'calcium carbonate' # Calcite  \_chemical\_formula.sum 'C1 Ca1 O3'  \_cell.length\_a 4.989  \_cell.length\_b 4.989  \_cell.length\_c 17.062  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 120  \_cell.volume 367.7795  \_cell.formula\_units\_Z 6  \_space\_group.name\_H-M\_alt 'R -3 c'  \_space\_group.name\_Hall '-R 3 2"c'  \_space\_group.IT\_number 167  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  36 1/3-y,2/3-x,1/6+z  loop\_  \_atom\_site.label  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  C1 0.00000 0.00000 0.25000  O1 0.25930 0.00000 0.25000  Ca1 0.00000 0.00000 0.00000  \_topol\_net.id 1  \_topol\_net.overall\_topology\_RCSR pcu-b  loop\_  \_topol\_node.id  \_topol\_node.label  1 ZA1 # CO3  2 ZB1 # Ca  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.symop\_id\_2  \_topol\_link.translation\_2  \_topol\_link.distance  \_topol\_link.type  1 1 2 20 [-1 -1 0] 3.2122 gl # (CO3)-Ca  loop\_  \_topol\_atom.id  \_topol\_atom.node\_id  \_topol\_atom.atom\_label  \_topol\_atom.element\_symbol  \_topol\_atom.symop\_id  1 1 C1 C 1  2 1 O1 O 2  3 1 O1 O 2  4 1 O1 O 3  5 2 Ca1 Ca 1  #End of data\_example\_3 |

|  |
| --- |
| *Example 4. Connectivity of an underlying net of the cuprite (Cu2O) crystal structure. Oxygen atoms coincide with the nodes, while copper atoms represent the links. There are two interpenetrating nets of the diamond topology* **dia***.*  #\#CIF\_2.0  data\_example\_4  \_database.code\_ICSD 63281  \_chemical.name\_systematic 'copper(I) oxide'  \_chemical\_formula.sum 'Cu2 O1'  \_cell.length\_a 4.267  \_cell.length\_b 4.267  \_cell.length\_c 4.267  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 77.69051  \_cell.formula\_units\_Z 2  \_space\_group.name\_H-M\_alt 'P n -3 m'  \_space\_group.name\_Hall '-P 4bc 2bc 3'  \_space\_group.IT\_number 224  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  48 1/2-z,y,1/2-x  loop\_  \_atom\_site.label  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  O1 0.25000 0.25000 0.25000  Cu1 0.00000 0.00000 0.00000  \_topol\_net.id 1  \_topol\_net.z\_number 2  \_topol\_net.overall\_topology\_RCSR 'dia'  \_topol\_node.id 1  \_topol\_node.net\_id 1  \_topol\_link.id 1  \_topol\_link.node\_id\_1 1  \_topol\_link.node\_id\_2 1  \_topol\_link.symop\_id\_2 13  \_topol\_link.type gl  loop\_  \_topol\_atom.id  \_topol\_atom.node\_id  \_topol\_atom.link\_id  \_topol\_atom.atom\_label  \_topol\_atom.element\_symbol  1 1 . O1 O  2 . 1 Cu1 Cu  #End of data\_example\_4 |

|  |
| --- |
| *Example 5. Description of three nets associated with the crystal structure of metal-organic framework MOF-5. The first net (Net\_1) corresponds to the initial frameworks; the other two nets correspond to underlying nets for two possible representations of the structure: as zinc tetranuclear complex groups connected by benzene rings (Net\_2) or as interconnected zinc, oxygen atoms and benzenedicarboxylato ligands (Net\_3). The TOPOL\_NODE loop contains nodes for all three nets. The TOPOL\_ATOM loop describes atoms of the initial framework belonging to different nodes and links of each of the three nets. Note that atom H1 is represented three times in TOPOL\_ATOM, because it needed for each of the three nets.*  #\#CIF\_2.0  data\_example\_5  \_database.code\_CSD SAHYIK  \_chemical\_formula.sum 'C24 H12 O13 Zn4'  \_cell.length\_a 25.669  \_cell.length\_b 25.669  \_cell.length\_c 25.669  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 16913.24  \_cell.formula\_units\_Z 8  \_space\_group.name\_H-M\_alt 'F m -3 m'  \_space\_group.name\_Hall '-F 4 2 3'  \_space\_group.IT\_number 225  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  192 1/2-z,1/2+y,-x  \_atom\_site.label  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  H1 0.30590 0.19410 0.04520  C1 0.25000 0.25000 0.11050  C2 0.28310 0.21690 0.02630  C3 0.25000 0.25000 0.05380  O1 .25000 0.25000 0.25000  O2 0.28180 0.21820 0.13390  Zn1 0.29350 0.20650 0.20650  loop\_  \_topol\_net.id  \_topol\_net.label  \_topol\_net.special\_details  \_topol\_net.overall\_topology\_TOPOS  1 Net\_1 'Atomic network' 'Unknown'  2 Net\_2 'Underlying net with zinc tetranuclear complex groups as nodes and benzene rings as links (cluster representation)' 'pcu'  3 Net\_3 'Underlying net with zinc, oxygen and benzenedicarboxylato ligands as 4-coordinate nodes (standard representation)' 'fff'  loop\_  \_topol\_node.id  \_topol\_node.net\_id  1 1 # H1  2 1 # C1  3 1 # C2  4 1 # C3  5 1 # O1  6 1 # O2  7 1 # Zn1  8 2 # C6O13Zn4  9 3 # C8H4O4  10 3 # O1  11 3 # Zn1  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.symop\_id\_2  \_topol\_link.distance  \_topol\_link.type  1 1 3 . 0.9594 v # H1-C2  2 2 6 37 1.3013 v # C1-O2  3 2 4 . 1.4554 ar # C1-C3  4 3 3 26 1.3502 ar # C2-C2  5 3 4 . 1.3936 ar # C2-C3  6 5 7 . 1.9340 v # O1-Zn1  7 6 7 . 1.9114 v # O2-Zn1  8 8 8 26 12.8345 gl # (C6O13Zn4)-(C6H4)-(C6O13Zn4)  9 9 11 5 5.5309 gl # (C8H4O4)-Zn1  10 10 11 . 1.9340 v # O1-Zn1  loop\_  \_topol\_atom.id  \_topol\_atom.atom\_label  \_topol\_atom.node\_id  \_topol\_atom.link\_id  \_topol\_atom.symop\_id  \_topol\_atom.element\_symbol  1 H1 1 . 1 H  2 C1 2 . 1 C  3 C2 3 . 1 C  4 C3 4 . 1 C  5 O1 5 . 1 O  6 O2 6 . 1 O  7 Zn1 7 . 1 Zn  10 C1 8 . 1 C  11 C1 8 . 55 C  12 O2 8 . 55 O  13 O2 8 . 80 O  14 O2 8 . 6 O  15 O2 8 . 59 O  16 O2 8 . 42 O  17 O2 8 . 78 O  18 C1 8 . 78 C  19 C1 8 . 5 C  20 Zn1 8 . 6 Zn  21 Zn1 8 . 55 Zn  22 C1 8 . 69 C  23 C1 8 . 6 C  24 O2 8 . 41 O  25 O2 8 . 69 O  26 O1 8 . 1 O  27 O2 8 . 5 O  28 O2 8 . 82 O  29 Zn1 8 . 5 Zn  30 Zn1 8 . 1 Zn  31 O2 8 . 1 O  32 O2 8 . 37 O  33 O2 9 . 14 O  34 O2 9 . 26 O  35 O2 9 . 1 O  36 O2 9 . 37 O  37 C1 9 . 14 C  38 C3 9 8 14 C  39 C2 9 8 14 C  40 H1 9 8 14 H  41 C2 9 8 37 C  42 H1 9 8 37 H  43 C2 9 8 26 C  44 H1 9 8 26 H  45 C2 9 8 1 C  46 H1 9 8 1 H  47 C3 9 8 1 C  48 C1 9 . 1 C  49 O1 10 . 1 O  50 Zn1 11 . 1 Zn  #End of data\_example\_5 |

|  |
| --- |
| *Example 6. The net associated with the hydrogen-bonded H2NCN molecules of cyanamide crystal structure. The TOPOL\_NET category is not needed for this simple description, and TOPOL\_NODE simplifies to just \_topol\_node.id. Note that \_topol\_link.order is included, indicating the bond order for the covalent bonds.*  data\_example\_6  \_database.code\_CSD CYANAM01  \_chemical.name\_systematic cyanamide  \_chemical\_formula.sum 'C1 H2 N2'  \_cell.length\_a 6.856  \_cell.length\_b 6.628  \_cell.length\_c 9.147  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 415.654  \_cell.formula\_units\_Z 8  \_space\_group.name\_H-M\_alt 'P b c a'  \_space\_group.name\_Hall '-P 2ac 2ab'  \_space\_group.IT\_number 61  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  2 1/2-x,-y,1/2+z  3 1/2+x,1/2-y,-z  4 -x,1/2+y,1/2-z  5 -x,-y,-z  6 1/2+x,y,1/2-z  7 1/2-x,1/2+y,z  8 x,1/2-y,1/2+z  loop\_  \_atom\_site.label  \_atom\_site.fract\_x  \_atom\_site.fract\_y  \_atom\_site.fract\_z  H1 0.19700 0.11100 0.02200  H2 0.05900 0.26600 0.08700  C1 0.14120 0.08370 0.23120  N1 0.14590 0.16990 0.10179  N2 0.14010 0.01250 0.34609  loop\_  \_topol\_node.id  1 2 3 4 5  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.distance  \_topol\_link.symop\_id\_2  \_topol\_link.translation\_2\_x  \_topol\_link.translation\_2\_y  \_topol\_link.translation\_2\_z  \_topol\_link.type  \_topol\_link.order  1 1 4 0.8988 1 0 0 0 v 1 # H1-N1  2 1 5 2.1228 2 0 0 -1 hb ? # H1-N2  3 2 4 0.8826 1 0 0 0 v 1 # H2-N1  4 2 5 2.2152 4 0 0 0 hb ? # H2-N2  5 3 5 1.1520 1 0 0 0 v 3 # C1-N2  6 3 4 1.3148 1 0 0 0 v 1 # C1-N1  loop\_  \_topol\_atom.id  \_topol\_atom.atom\_label  \_topol\_atom.node\_id  \_topol\_atom.element\_symbol  1 H1 1 H  2 H2 2 H  3 C1 3 C  4 N1 4 N  5 N2 5 N  #End of data\_example\_6 |

|  |
| --- |
| *Example 7. The underlying net of the FAU zeolite framework of faujasite (representing all aluminum and silicon atoms as Si, and ignoring all other elements). The tiling is identified by its signature, 2[4^6.6^2]+[4^6.6^8]+[4^18.6^4.12^4].*  #\#CIF\_2.0  data\_example\_7 # FAU  \_cell.length\_a 24.345  \_cell.length\_b 24.345  \_cell.length\_c 24.345  \_cell.angle\_alpha 90  \_cell.angle\_beta 90  \_cell.angle\_gamma 90  \_cell.volume 14428.77  \_cell.formula\_units\_Z 192  \_space\_group.name\_H-M\_alt 'F d -3 m'  \_space\_group.name\_Hall '-F 4vw 2vw 3'  \_space\_group.IT\_number 227  loop\_  \_space\_group\_symop.id  \_space\_group\_symop.operation\_xyz  1 x,y,z  …  192 3/4-z,1/2+y,1/4-x  \_atom\_site.label Si1  \_atom\_site.fract\_x 0.94690  \_atom\_site.fract\_y 0.12510  \_atom\_site.fract\_z 0.03640  \_topol\_net.id 1  \_topol\_net.overall\_topology\_iza FAU  \_topol\_node.id 1  \_topol\_node.net\_id 1  \_topol\_node.label Si  loop\_  \_topol\_link.id  \_topol\_link.node\_id\_1  \_topol\_link.node\_id\_2  \_topol\_link.distance  \_topol\_link.symop\_id\_2  \_topol\_link.translation\_2  \_topol\_link.type  1 1 1 3.0470 47 [0 0 0] gl  2 1 1 3.0473 13 [1 1 0] gl  3 1 1 3.0539 41 [0 0 0] gl  4 1 1 3.0814 42 [1 0 -1] gl  \_topol\_atom.id 1  \_topol\_atom.node\_id 1  \_topol\_atom.atom\_label Si1  \_topol\_atom.element\_symbol Si  \_topol\_tiling.id 1  \_topol\_tiling.net\_id 1  \_topol\_tiling.signature '2[4^6.6^2]+[4^6.6^8]+[4^18.6^4.12^4]'  \_topol\_tiling.vertices 1  \_topol\_tiling.edges 4  \_topol\_tiling.faces 5  \_topol\_tiling.tiles 3  \_topol\_tiling.d\_size 24  #End of data\_example\_7 |

**4. Available reference implementations**

The CIF topological data can be read and generated by the *ToposPro* program package, version 5.4 or higher (https://topospro.com) (Blatov *et al*., 2014) and online service *TopCryst* (https://topcryst.com). *Jmol*/*JSmol* software (<https://jmol.sourceforge.net>, <https://github.com/BobHanson/Jmol-SwingJS>) can also read and visualize the nets specified in the topoCIF format. The topological databases Reticular Chemistry Structure Resource (RCSR, http://rcsr.net/) and ToposPro Collections (TTD and TTO, https://topospro.com) use the topoCIF topological descriptors. Thus, one can use the topoCIF format for exporting or importing information to or from these databases.

The periodic net described by a topoCIF file is fully defined by its set of TOPOL\_ATOM, TOPOL\_NODE, and TOPOL\_LINK data. The method of constructing a net for visualization or analysis from this data is not trivial, since generally a CIF file only provides locations for the asymmetric unit. Below we describe a sketch of the sequence of steps used by Jmol. The presumption is that we start with a request for the visualization of a net based on a given set of unit cells or a specified volume of space to be filled. Net construction proceeds in the following order:

1. Read relevant \_cell, \_space\_group, \_space\_group\_symop, \_atom\_site, \_topol\_net, \_topol\_link, \_topol\_node, and \_topol\_atom data (which could come in any order within a CIF file) from a data block.
2. Process all the TOPOL\_ATOM items by
   1. applying indicated symmetry operations and translations, and
   2. associating each atom with its designated link or node.
3. Process all additional link-specific end point symmetry operations and translations.
4. Apply all space group symmetry operations for all links and nodes, taking care not to duplicate atom positions within a given net.
5. Remove any unconnected or unassociated nodes that are present within the specified space.

A reference dual Java/JavaScript implementation of Jmol/JSmol can be found at SourceForge (<https://sourceforge.net/p/jmol/code/HEAD/tree/trunk/Jmol/src/org/jmol/adapter/readers/cif/TopoCifParser.java>) or GitHub (<https://github.com/BobHanson/Jmol-SwingJS/blob/master/src/org/jmol/adapter/readers/cif/TopoCifParser.java>).

**References**

Alexandrov, E. V., Blatov, V. A., Kochetkov, A. V. & Proserpio, D. M. (2011). *Underlying nets in three-periodic coordination polymers: topology, taxonomy and prediction from a computer-aided analysis of the Cambridge Structural Database. CrystEngComm*, **13**, 3947-3958.

Aman, F., Asiri, A. M., Siddiqui, W. A., Arshad, M. N., Ashraf, A., Zakharov, N. S. & Blatov, V. A. (2014). *Multilevel topological description of molecular packings in 1,2-benzothiazines. CrystEngComm*, **16**, 1963-1970.

Barthel, S., Alexandrov, E. V., Proserpio, D. M., Smit, B. (2018). *Distinguishing Metal-Organic Frameworks. Cryst. Growth Des.*, **18**, 1738–1747.

Blatov, V.A. (2004). *Voronoi-Dirichlet polyhedra in crystal chemistry: theory and applications. Cryst. Rev.* **10**, 249-318.

Blatov, V. A. (2007). *Topological relations between three-dimensional periodic nets. I. Uninodal nets. Acta Cryst.* **A63**, 329–343.

Blatov, V. A., Delgado-Friedrichs, O., O’Keeffe, M. & Proserpio, D. M. (2007). *Three-periodic nets and tilings: natural tilings for nets. Acta Cryst*. **A63**, 418-425.

Blatov, V.A., O’Keeffe, M. & Proserpio, D. M. (2010). *Vertex-, face-, point-, Schläfli-, and Delaney-symbols in nets, polyhedra and tilings: recommended terminology*. *CrystEngComm*, **12**, 44-48.

Blatov, V. A., & Proserpio, D. M. (2009). *Topological relations between three-dimensional periodic nets. I. Binodal nets. Acta Cryst.* **A65**, 202-212.

Blatov, V.A., Shevchenko, A.P. & Proserpio, D.M. (2014). *Applied topological analysis of crystal structures with the program package ToposPro. Cryst. Growth Des.* **14**, 3576-3586.

Bonneau, C., O’Keeffe, M., Proserpio, D. M., Blatov, V. A., Batten, S. R., Bourne, S. A., Lah, M. S., Eon, J.-G., Hyde, S. T., Wiggin, S. B. & Öhrström, L. (2018) *Deconstruction of Crystalline Networks into Underlying Nets: Relevance for Terminology Guidelines and Crystallographic Databases. Cryst. Growth Des.*, **18**, 3411–3418

Delgado-Friedrichs, O. & O’Keeffe, M. (2003). *Identification of and symmetry computation for crystal nets. Acta Cryst.* **A59**, 351-360.

Delgado-Friedrichs, O. & O’Keeffe, M. (2005). *Crystal nets as graphs: terminology and definitions. J. Solid State Chem*. **178**, 2480-2485.

Delgado-Friedrichs, O., Foster, M. D., O'Keeffe, M., Proserpio, D. M., Treacy, M. M. J. & Yaghi, O. M. (2005). *What do we know about nets? J. Solid State Chem*. **178**, 2533-2554.

Eon, J-.G (2016) *Topological features in crystal structures: a quotient graph assisted analysis of underlying nets and their embeddings.* *Acta Cryst.* **A72**, 268-293.

Francl, M. (2009) *Stretching topology. Nat. Chem.*, **1**, 334-335.

O'Keeffe, M., Peskov, M. A., Ramsden, S. J. & Yaghi, O. M. (2008). *The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. Acc. Chem. Res.* **41**, 1782-1789.

Koch, E., Fischer, W. & Sowa, H. (2006). *Interpenetration of homogeneous sphere packings and of two-periodic layers of spheres. Acta Cryst.* **A62**, 152-167.

Ramsden, S. J., Robins, V., Hyde, S. T. (2009). *Three-dimensional Euclidean nets from two-dimensional hyperbolic tilings: kaleidoscopic examples. Acta Cryst.* **A65**, 81–108.

Shevchenko, A.P. & Blatov, V.A. (2021). *Simplify to understand: how to elucidate crystal structures? Struct. Chem.* **32**, 507-519.