Inorganic Crystal Structures

DOI: 10.1002/anie.201304374

Which Inorganic Structures are the Most Complex?

*Sergey V. Krivovichev**

crystallography · information theory · inorganic structures · structural complexity · structural elucidation

The discovery of the diffraction of X-rays on crystals opened up a new era in our understanding of nature, leading to a multitude of striking discoveries about the structures and functions of matter on the atomic and molecular scales. Over the last hundred years, about 150000 of inorganic crystal structures have been elucidated and visualized. The advent of new technologies, such as area detectors and synchrotron radiation, led to the solution of structures of unprecedented complexity. However, the very notion of structural complexity of crystals still lacks an unambiguous quantitative definition. In this Minireview we use information theory to characterize complexity of inorganic structures in terms of their information content.

1. Introduction

The evolution of matter towards states of higher complexity in the course of the evolution of the Universe is one of the most interesting topics in modern science and in modern chemistry in particular.[1] A number of different approaches exist that allow the complexity of chemical, biological, technological, and even social systems to be quantitatively evaluated.[2] Most of these approaches are based upon the extensive use of concepts and techniques borrowed from the information theory developed in the second half of the 20th century as a theory of communication and transmission of information.[3] Nowadays complexity theory is a well-developed field of computer science with interdisciplinary applications in many areas of theoretical and applied sciences.[4] It is interesting, however, that complexity theorists have always regarded crystals as "intuitively simple objects",[5] in contrast to much more complex and dynamic biological systems. On the other hand, it was largely appreciated that crystal structures may possess different degrees of complexity, which has been reflected in the titles of scientific papers with the use of expressions such as "complicated structure", "the most complex structure", "complexity", "complex superstructures", "a masterpiece of structural complexity".[6] However, until recently, the complexity of crystal structures was destined to be "largely a qualitative, frequently intuitive, notion",[7] which escaped a quantitative definition.

Perhaps, the first attempt to evaluate structural complexity of crystals should be assigned to the Linus Pauling's fifth ("parsimony") rule "determining the structure of complex ionic crystals":[8] "the number of different types of constituents in a crystal tends to be small." Hawthorne[9] noted that "this rule has not seen much use, apart from in teaching, but it certainly seems to hold reasonably well, and its underlying cause is certainly of interest."

The certain advantage of crystals is that their almost infinite network of atoms and bonds can be described by a rather small volume of space called the elementary unit cell. Of even smaller volume is the asymmetric unit, which only contains symmetrically inequivalent atoms. To provide a quantitative definition of crystal-structure complexity, Baur et al.[10] suggested to use topological and crystallographic parsimony indices, I_t and I_c, defined by Equations (1) and (2),

$$I_t = (t-e)/t \quad (1)$$

$$I_c = (c-e)/c \quad (2)$$

where e is the number of different chemical elements, and t and c are the numbers of topologically and crystallographically distinct atom sites in the asymmetric unit, respectively. Baur et al.[10] pointed out that the topological parsimony index is of primary importance and suggested subdividing structures into lavish ($I_t > 0.66$), intermediate ($0.66 > I_t > 0.33$), and parsimonius ($I_t < 0.33$). Both indices are not sensitive either to chemical complexity or symmetry and reflect only relations between the number of chemical elements and the number of topologically or crystallographically distinct atom sites.

[*] Prof. Dr. S. V. Krivovichev
Department of Crystallography, St. Petersburg State University
University Emb. 7/9, 199034 St. Petersburg (Russia)
E-mail: s.krivovichev@spbu.ru

An alternative approach to structural complexity was suggested by Lister et al.,[6b,c] who used the number of atoms in the asymmetric unit as a measure of complexity for a commensurate structure. This method has many advantages but, unfortunately, does not provide an appropriate quantification of the complexity of high-symmetry structures, such as, paulingite, a complex natural zeolite, which was identified by Mackay and Klinowski[11] as one of the most complex inorganic structure.

Mackay[12] and later Estevez-Rams and González-Férez[13] proposed to estimate the complexity of crystal structures on the basis of the complexity of algorithms that can be used to generate them from an unordered array of atoms. This concept is of great interest, but, in its present state, requires a high level of abstraction and imagination and cannot be applied to crystal structures in an easy and unambiguous way.

In this Minireview, a different method is employed to evaluate the complexity of structures, which is based upon the relatively simple application of Shannon information theory and allows both size- and symmetry-dependent properties of complex structures to be quantitatively evaluated, thus combining their symbolic and combinatorial complexities[14] in the same measure.

2. Basic Concepts

In the framework of classical information theory, information content H of a message M consisting of v symbols that belong to the k equivalence classes is calculated according to the following Shannon formula [Eq. (3)]

$$H = -\sum_{i=1}^{k} p_i \log_2 p_i \ \text{[bits/symbol]} \tag{3}$$

where p_i is the probability of occurrence of the ith symbol [Eq. (4)]:

$$p_i = m_i/v \tag{4}$$

where m_i is the number of symbols in M that belong to the ith equivalence class.

Following theoretical works on the topological complexity of graphs and networks,[2b,15] Krivovichev[16] proposed to consider the reduced elementary unit cell of a crystal structure as a message consisting of v atoms. The atoms are subdivided into equivalence classes according to their crystallographic orbits, that is, two atoms belong to the same class if they are symmetrically equivalent. This led to the following formula to calculate the structural information content, I_G, of a crystal structure [Eq. (5)]

$$I_G = -\sum_{i=1}^{k} p_i \log_2 p_i \ \text{[bits/atom]} \tag{5}$$

where k is the number of crystallographic orbits and p_i is the probability of occurrence of the atom of the ith crystallographic orbit [Eq. (6)]

$$p_i = m_i/v \tag{6}$$

where m_i is the multiplicity of the crystallographic orbit relative to the reduced unit cell, and v is the number of atoms in the reduced unit-cell.

For instance, the structure of lopezite, $K_2Cr_2O_7$,[17] is triclinic, space group $P\bar{1}$. All the atoms in the structure occupy $2i$ Wyckoff sites: there are four K, four Cr, and fourteen O sites, which corresponds to the total number of 22 equivalence classes. The total number of atoms in the reduced unit cell (which in this case coincides with the crystallographic unit cell) is 44. Atoms from each equivalence class has the same probability of occurrence, which is equal to $2/44 = 0.0455$. The structural information content for lopezite is given by Equation (7).

$$I_G = -22 \times (2/44) \log_2 2/44 = 4.459 \ \text{bits/atom} \tag{7}$$

The total information content of a crystal structure is determined by the Equation (8).

$$I_{G,\text{total}} = v \times I_G \ \text{(bits/unit-cell)} \tag{8}$$

For lopezite, $I_{G,\text{total}} = 44 \times 4.459 = 196.196$ bits per unit-cell (bits/u.c.).

The maximal structural information content, $I_{G,\max}$, for a crystal structure with v atoms occurs, where all atoms are non-equivalent according to the corresponding space group. It can be calculated as Equation (9).

$$I_{G,\max} = \log_2 v \ \text{(bits/atom)} \tag{9}$$

For I_G to be independent of v, we may define the normalized structural information content, $I_{G,\text{norm}}$, as Equation (10).

$$I_{G,\text{norm}} = I_G/I_{G,\max} = -[\sum_{i=1}^{k} p_i \log_2 p_i]/\log_2 v \tag{10}$$

For the structure of lopezite, $I_{G,\max} = \log_2 44 = 5.459$ bits/atom and $I_{G,\text{norm}} = 4.459/5.459 = 0.817$. Note that $I_{G,\text{norm}}$ is a dimensionless quantity in the range from 0 to 1.

The information-based complexity measures allow us to introduce the concept of information density, which can be defined as a number of bits of information per cubic Ångström [Eq (11)].

Sergey Krivovichev received his Ph.D. (1997) and D.Sc. (2002) degrees from St. Petersburg State University and carried our post-doctoral research in the U.S., Germany, Switzerland, and Austria. Since 2005, he is a Full Professor and Chairman of Department of Crystallography, St. Petersburg State University. His research interests include inorganic structural chemistry, mineralogy, crystallography, nuclear materials, and applications of complexity theory to natural and artificial objects.

$$\rho_{\inf} = I_{G,\text{total}}/V \ (\text{bits Å}^{-3}) \quad (11)$$

For lopezite, $\rho_{\inf} = 0.270$ bits Å^{-3}. For another modification of $K_2Cr_2O_7$,[18] $\rho_{\inf} = 0.156$ bits Å^{-3}, which indicates its lower complexity compared to lopezite. It is important that information density should be calculated using the volume of the reduced unit cell.

For two structures with similar unit cells but different symmetry, information density will be higher for the structure with lower symmetry (and higher complexity). For instance, cation ordering within the same unit cell would lead to the formation of a superstructure with higher information density than that of the original structure. This situation shows that unit-cell volume alone cannot be considered as a measure of structural complexity, since it reflects only the symbolic complexity of the system and is not sensitive to its combinatorial complexity.

Information density can also be used as a measure of complexity for the crystal structures of solids with similar composition. The fact that complexity is a function of both information content and unit-cell volume makes it dependent upon the size of atoms and ions that make up the respective crystal, as well as upon temperature and pressure. Suanite, $Mg_2B_2O_5$,[19] and natrosilite, $Na_2Si_2O_5$,[20] have the same total structural information content (114.117 bits/u.c.), but their information densities are drastically different (0.333 and 0.242 bits Å^{-3}, respectively), which is a consequence of the different ionic radii of the Mg^{2+} and Na^+ ions, on the one hand, and of the B^{3+} and Si^{4+} ions, on the other. However, while information density cannot be used to compare structures with different chemical compositions, it is a useful parameter for the investigation of behavior of structural information in the course of phase transitions induced by temperature and pressure, that is, when chemical composition is a constant parameter.

Table 1 presents comparison of the values of the complexity parameters defined according to Baur et al.,[10] Lister et al.,[6b,c] and this work, applied to different simple and complex inorganic structures. These data indicate certain advantages in the use of information-based complexity parameters (and, in particular, the $I_{G,\text{total}}$ parameter) as compared to the purely crystallographic parameters (v, c, I_t, and I_c). For instance, according to the intuitive feelings, $BaTiO_3$ is more complex than rutile, TiO_2 (rutile), whereas rutile is more complex than NaCl (halite). Both parsimony indices, I_t and I_c,[10] cannot discriminate between the complexities of the three structures, whereas the number of atoms within the asymmetric unit (c)[6b,c] cannot distinguish between the complexities of tetragonal rutile and cubic halite. On the other hand, ranking of the structures according to the $I_{G,\text{total}}$ values is in agreement with the intuitive feelings (in bits per unit cell): 6.855 ($BaTiO_3$) > 5.510 (rutile) > 2.000 (halite). In general, it may be concluded that the parsimony indices[10] do not possess enough discriminative power, whereas the c parameter underestimates complexities of highly symmetrical structures and structures with non-general orbits.

Since total information content of the crystal structure, $I_{G,total}$, is defined per unit cell, it is possible to calculate a total information amount of a particular crystal. For instance, a 1 mole crystal of α-quartz, space group $P322_1$, $Z = 3$, contains approximately 2.00738×10^{23} unit cells calculated as the Avogadro number, 6.02214×10^{23},[21] divided by $Z = 3$. The total information content of the crystal structure of α-quartz is $I_{G,\text{total}} = 8.265$ bits/u.c. Therefore, the total information content of the 1 mole quartz crystal is equal to approximately 16.59100×10^{23} bits. It should be emphasized that, clearly, structural information of a crystal consisting of billions of unit cells is redundant or algorithmically compressible.[12] Nevertheless, the full amount of this (redundant) information can be calculated from the amount of the respective crystalline substance expressed either in molar, mass, or volumnar quantities. This means that, similar to mass, entropy, volume, for example, information is an "extensive" property as opposed to the "intensive" properties such as temperature and pressure.

To explain this point further, consider the following sequences: a) 010101010101010101, b) 101001101001101001, and c) 010001111000101101. To store these sequences, any computer would need exactly 18 bits for each, which means that the amount of information expressed in bits is identical for the three sequences. However, the sequence (c) cannot be compressed, whereas the sequences (a) and (b) are algorithmically compressible: the sequence (a) can be represented as "repeat 01 nine times", whereas (b) as "repeat 101001 three times". By analogy, crystals can be constructed by repeating its unit cell in a three-dimensional space; however, its full information content will be a function of the number of these repeats, just as the full information content of the sequence (b) is the function of the number of repeats of the sequence "101001". It can be argued that, in the framework of this approach, a piece of glass would contain more information than the crystal with the corresponding composition, since, because of the absence of the long-range order, the structure of the glass is unique in each point of the piece. However, this information is meaningless, since local distortion of the struc-

Table 1: Comparison of the values of different complexity parameters for the same inorganic crystal structures

Structure	ICSD code	v [atoms][a]	c [atoms][b]	I_t[c]	I_c[d]	I_G [bits]	$I_{G,\text{total}}$ [bits/u.c.]	$I_{G,\text{norm}}$
NaCl	53815	2	2	0	0	1.000	2.000	1.000
TiO_2 rutile	53997	6	2	0	0	0.918	5.510	0.355
$BaTiO_3$ cubic	56093	5	3	0	0	1.371	6.855	0.590
SiO_2 coesite	36261	24	7	0	0.71	2.752	66.039	0.600
$Ba_2Ti_9O_{20}$	35711	124	62	0.90	0.95	5.954	738.320	0.856
$Ba_6Ti_{17}O_{40}$	49576	126	34	0.89	0.91	5.057	637.137	0.725
SiP_2O_7 cubic	66861	1080	50	0	0.94	5.570	6015.280	0.553
$Cu_2Li_2P_6O_{18}$	2149	56	28	0.73	0.86	4.807	269.212	0.828

[a] Number of atoms per reduced unit cell. [b] Number of atoms per asymmetric unit.[6b,c] [c] Topological parsimony index.[10] [d] Crystallographic parsimony index.[10]

ture of the glass would not alter its properties.

The I_G value proposed above measures structural information per atom, whereas the $I_{G,\text{total}}$ value defines information amount per smallest unit, which forms the whole structure by its repeating in three dimensions. As can be seen from Table 1, the most adequate parameter to evaluate complexity of a crystal structure is $I_{G,\text{total}}$.

3. Complexity of Inorganic Structures: Statistics

To analyze the complexity of inorganic structures on a large-scale basis, complexity parameters have been calculated for all the crystal-structure entries stored in the Inorganic Crystal Structure Database (ICSD)[22] using the TOPOS program package.[23] Figure 1 shows distribution of

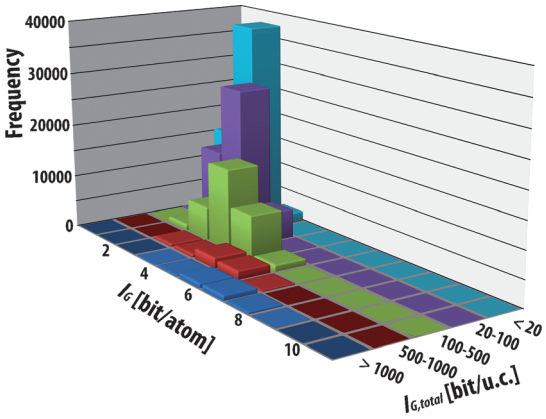

Figure 1. Histogram showing distribution of inorganic crystal-structure reports stored in the Inorganic Crystal Structure Database (ICSD) according to the values of the information-based I_G and $I_{G,\text{total}}$ parameters.

the ICSD reports according to the information contents per atom (I_G) and per unit-cell ($I_{G,\text{total}}$). According to the $I_{G,\text{total}}$ values, all structures can be subdivided into very simple, simple, intermediate, complex, and very complex (Table 2). It can be seen from Figure 1 that most ICSD entries are for very simple and simple structures that together constitute about 75% of the ICSD data. This reflects the fact that structurally and chemically simple inorganic compounds have been most studied over the last hundred years, though the general current trend is towards the elucidation of structures with increasing complexity (see below).

Table 2: Classification of inorganic crystal structures according to their complexity (information content per unit cell).

Category	$I_{G,\text{total}}$ [bits/u.c.]	Number of ICSD entries
very simple	< 20	ca. 55 000
simple	20–100	ca. 50 000
intermediate	100–500	ca. 33 000
complex	500–1000	ca. 5000
very complex	> 1000	ca. 3000

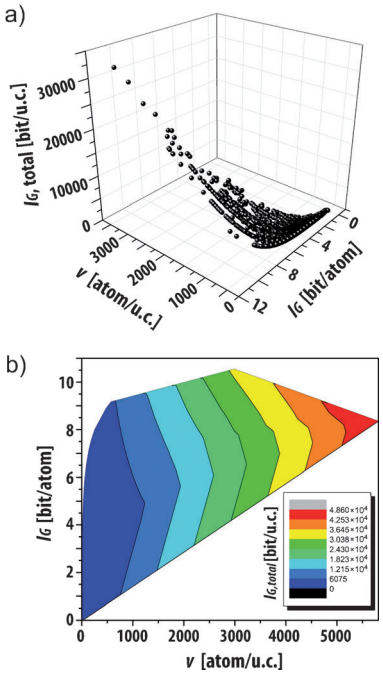

Figure 2. Diagrams showing relations between complexity parameters for the ICSD reports: number of atoms per unit cell (v), structural information content per atom (I_G) and per unit cell ($I_{G,\text{total}}$): point diagram (a); each point corresponds to a single ICSD report (two reports with $I_{G,\text{total}} > 35\,000$ bits are omitted for clarity) and a contour map (b) colors show the range of the $I_{G,\text{total}}$ values).

Figure 2 shows relations between the information contents per atom (I_G) and per unit-cell ($I_{G,\text{total}}$), and the number of atoms in the unit cell (v) (note that all the quantities are referred to the reduced unit cells). It can be seen that there is a general positive correlation between v and $I_{G,\text{total}}$. However, structures with the same number of atoms per unit cell may differ in their $I_{G,\text{total}}$ values by several hundreds of bits, which shows that, in agreement with the previous conclusion, the purely crystallographic v parameter cannot be used as a single complexity measure, since it accounts for the size of the structure only, and not for its symmetry. The symmetry-sensitive parameter is I_G, and the $I_{G,\text{total}}$ parameter provides combination of both size- and symmetry-sensitive complexity measures.

4. Most Complex Inorganic Structures

Analysis of the 2000 most-complex structures stored in the ICSD according to their belonging to specific classes of inorganic compounds is shown in Figure 3. About 90% of all these compounds are structures based upon nanoscale-sized clusters with 82% represented by Mo-, W-, and V-based polyoxometallates. Among other classes are: oxides and oxysalts (e.g., silicates, sulfates, borates), zeolites and microporous framework materials, intermetallic compounds, borides and silicides, superstructures of simpler structure types, sulfides and selenides, and halogenides.

The most complex inorganic structure reported to date is that of the intermetallic compound $Al_{55.4}Cu_{5.4}Ta_{39.1}$ (ACT-

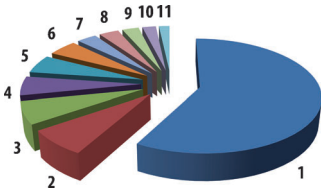

Figure 3. Classification of the 2000 most-complex inorganic structures according to their chemical identity: 1) Mo-, W-, V-based polyoxometallates; 2) oxides and oxysalts; 3) structures containing (nano)clusters; 4) zeolites and microporous framework compounds; 5) intermetallic compounds (including borides and silicides); 6) fullerene-based structures; 7) uranyl peroxide nanospheres; 8) superstructures of simple structure types; 9) Nb-, Ti-based polyoxometallates; 10) sulfides and selenides; 11) halogenides.

71).[24] Its total structural information content is equal to 48538.637 bits/u.c. The reduced unit cell contains $v = 5814$ atom sites and the structural information content per atom (I_G) is 8.349 bits. The exceptional complexity of this structure is the result of the combination of different structural modules, which was described by Steurer and coworkers[24] as follows: The $Al_{12}Ta_{28}$ fullerene cluster shells (orange polyhedra in Figure 4a consisting of 40 faces and 76 vertices) are linked by the Ta_{15} bifrusta (red polyhedra consisting of 12 faces and 15 vertices) into a porous framework with large cavities occupied by supertetrahedral clusters of two types. The first type (Figure 4b,c) consists of 146 Ta-centered Al_{12} Friauf polyhedra, Ta_8 hexagonal bipyramids, and $Al_{12}Ta_{28}$ fullerene shells. The second type is formed by ten $Al_{102}Ta_{57}$ fullerenes (Figure 4e). The supertetrahedral clusters

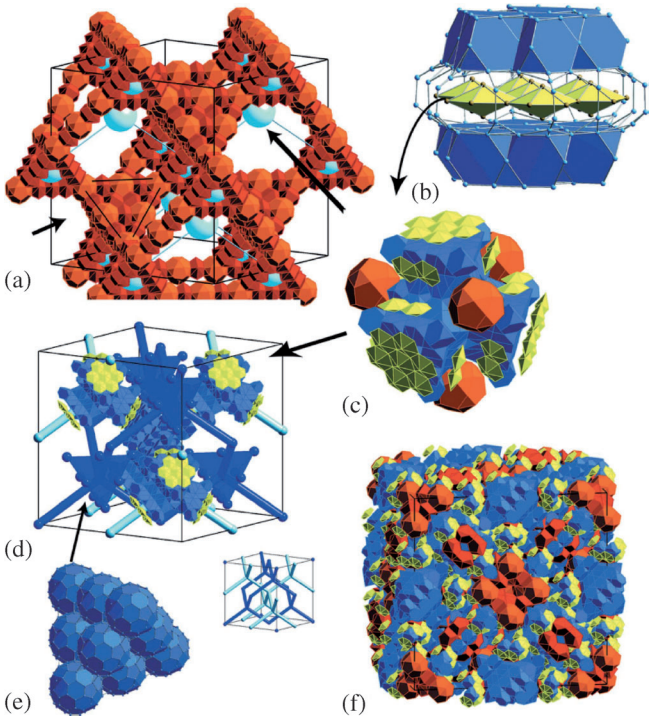

Figure 4. Structural building principle of $Al_{55.4}Cu_{5.4}Ta_{39.1}$ (ACT-71). Reproduced with permission from Ref. [24].

form two interpenetrating diamond-type networks (Figure 4d) and, together with the fullerene–bifrustum framework (Figure 4a), form the unit cell of extraordinary complexity (Figure 4f).

The second most complex structure is that of $(C(NH_2)_3)_{136}[\{(C(NH_2)_3)_{20}(H_2O)_{100}\}\subset\{(Mo)Mo_5O_{21}(H_2O)_6\}_{12}\{Mo_2O_4(H_2PO_2)\}_{20}\{Mo_2O_4(SO_4)\}_{10}][\{(C(NH_2)_3)_{20}(H_2O)_{80}\}\subset\{(Mo)Mo_5O_{21}(H_2O)_6\}_{12}\{Mo_2O_4(H_2PO_2)\}_{12}\{Mo_2O_4(SO_4)\}_{18}](H_2O)_{400}$,[25] which contains spherical polyoxomolybdate cluster of the type $\{(Mo)Mo_5\}_{12}(spacer)_{30}$ (one of the "Keplerates")[26] (Figure 5a,b). The nanocavity inside the cluster is occupied by the $(H_2O)_{100}$ 'nanodrop' of water with H_2O

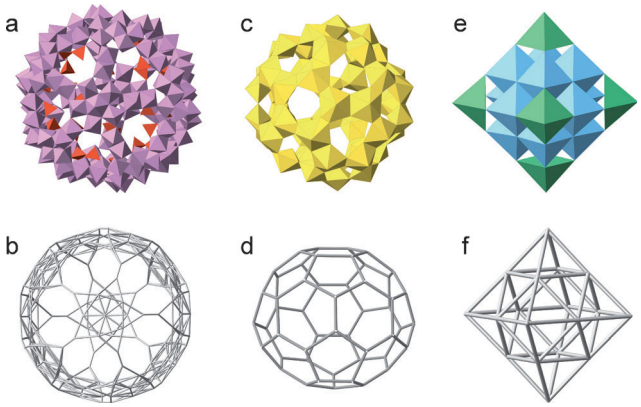

Figure 5. Polyoxometalates nanoscale clusters from selected very complex inorganic structures (skeletal representation is given for each cluster showing the topology of the metal-O-metal links): a,b) "Kepplerate" cluster from the structure of $\{[(NH_2)_3C]_{52}\}\{(Mo)Mo_5O_{21}(H_2O)_6\}_{12}\{Mo_2O_4(SO_4)\}_{10}\{Mo_2O_4(H_2PO_2)\}_{20}$ (pink and red polyhedra are Mo and P coordination polyhedra, respectively); c,d) uranyl peroxide nanosphere with the C_{50} fullerene topology (yellow polyhedra are uranyl hexagonal bipyramids); e,f) 18-nuclear Ti-Nb-polyoxometalate cluster from the structure of $[N(CH_3)_4]_{20}[Ti_{12}Nb_6O_{44}]_2(H_2O)_{73}$ (Ti and Nb polyhedra are shown as blue and green, respectively).

molecules arranged in three concentric shells. The enormous complexity ($I_{G,total} = 45341.050$ bits/u.c., $v = 5128$ atoms/u.c.) is not only due to the size of the clusters, but also due to the fact that there are two crystallographically independent clusters in the structure with $\bar{3}m$ and $2/m$ site symmetries.

Very complex structures with giant clusters are dominated by Mo-, W-, and V-polyoxometallates. However, oxo-nanoclusters of other metals can also assemble to form structures of high complexity. A prominent example is the uranyl peroxide nanospheres formed by linkage of uranyl ions, UO_2^{2+}, through peroxide groups (Figure 5c,d).[27] The maximal number of uranyl ions in such clusters prepared to date is 120, and some of the clusters adopt fullerene topologies.[28] It is noteworthy that uranyl peroxide nanospheres are porous and their cavities may accommodate various cations and H_2O molecules. Figure 5e shows the $[Ti_{12}Nb_6O_{44}]^{10-}$ cluster from the structure of $[N(CH_3)_4]_{20}[Ti_{12}Nb_6O_{44}]_2(H_2O)_{73}$.[29] Despite the obvious high symmetry of the cluster ($m\bar{3}m$) that can be appreciated from its skeletal representation (Figure 5f), the symmetry of the crystal structure is monoclinic (space group $P2_1/c$).

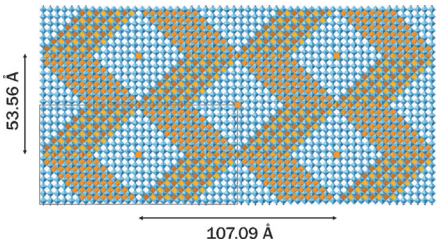

Figure 6. The nanoscale-domain superstructure of $Nd_{7/12}Li_{1/4}TiO_3$ consisting of square-shaped Li-rich regions (white) separated by zig-zagging boundary regions of a Nd-rich structure (orange).

Another complexity-generating mechanism in inorganic materials is formation of superstructures either by atomic ordering or displacive phase transitions. Figure 6 shows an ordered nano-checkerboard superlattice observed in the structure of $Nd_{7/12}Li_{1/4}TiO_3$ by transmission electron microscopy and modeled by Rietveld refinement.[30] The structure in general belongs to the perovskite structure type (ABO_3) with Ti in the *B* site and Li and Nd in the *A* site. The Li–Nd ordering results in formation of square-shaped nanoscale domains of a Li-rich phase separated by a zig-zagging boundary region consisting of a Nd-rich phase.[30] The resulting supercell has the dimensions $14a_p \times 28a_p \times 2a_p$, where a_p is the unit-cell parameter of a simple perovskite. The total information content of the structure equals to 16634.804 bits/u.c. with $v = 1973$ atoms/u.c.

An example of a very complex structure with complexity generated by displacive phase transitions on cooling is that of $[Mo_2P_4O_{15}]$ which can be written as $[(MoO)_2P_4O_{13}]$.[6b,c] The structure consists of corner-sharing MoO_6 octahedra interlinked by linear P_4O_{13} tetramers of tetrahedra. Its topological complexity (i.e., complexity of an archetype framework) is rather small ($I_{G,\text{total}} = 144.477$ bits/u.c.),[31] whereas real structures are very complex (7748.048 bits/u.c. for monoclinic 120 K α-phase and 4022.549 bits/u.c. for triclinic 573 K β-phase).[6b,c] The high complexity of the real structures compared to the much simpler archetype is due to the local bonding requirements that led to the formation of extended superstructures. This example is remarkable in that the symmetry of a high-temperature β-phase (space group $P\bar{1}$) is lower than the symmetry of a low-temperature α-phase (space group *Pn*). However, the information amount of the high-temperature phase is smaller than that of the low-temperature phase, which shows that, at least in this case, the information amount decreases with the increasing temperature, despite the decreasing symmetry.

The high topological and structural complexity of zeolites and microporous oxide-based frameworks has long been recognized. One of the most complex zeolites[16c] and the most complex mineral known to date[16d] is paulingite, $K_6Ca_{16}(Al_{38}Si_{130}O_{336})(H_2O)_{113}$,[32] with $I_{G,\text{total}} = 6766.998$ bits/u.c. Its structure (Figure 7a,b) is a complex aluminosilicate framework of three different types of secondary building units (Figure 7c) linked into two interpenetrating **pcu** (primitive cubic) nets (Figure 7d,e) connected by additional T nodes. The most reasonable explanation of zeolite complexity is their growth by self-assembly of prenucleation nanoclusters

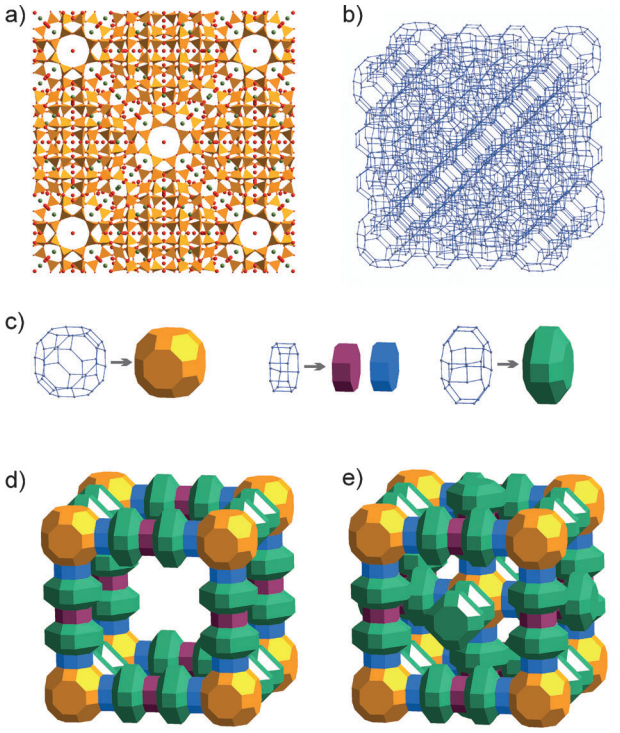

Figure 7. The structure of paulingite viewed along the *a* axis (a) and the nodal representation of its tetrahedral framework (b). The framework can be considered as consisting of cages (c) that share edges to form a three-dimensional framework (d). Two frameworks interpenetrate and are linked by additional nodes (not shown) to form a complete paulingite topology (e).

already existing in the crystallization media.[33] Among the zeolite framework topologies known to date,[34] the paulingite PAU framework is the second in its topological complexity (4763.456 bits/u.c.), after the SFV framework reported for the synthetic zeolite SSZ-57 (19557.629 bits/u.c.).[35]

In general, oxides and oxysalts comprise about 6% of 2000 most complex inorganic structures (Figure 3). Borates are characterized by the intrinsic structural complexity, owing to the ability of the B^{3+} ion to occur in triangular and tetrahedral coordination environments in the same structure. Figure 8a shows the disposition of the $[B_7O_{12}]^{3-}$ borate layers in the structure of $Cs_3B_7O_{12}$.[36] Each layer consists of three sublayers. The sublayers **1** and **2** are identical and rather simple, whereas the sublayer **3** that links sublayers **1** and **2** contains triborate moieties of three geometrical types (Figure 8b), which leads to the occurrence of the very long unit-cell parameter of 59.911 Å. The resulting structure is very complex and its total information content equals to 6062.598 bits/u.c.

The structure of $Ba_{12}(SiO_3)_6(CO_3)_5Cl_2$[37] is an interesting example of the case when structural complexity is generated by the complex chemical composition. It can be described as being based upon rods of face-sharing ($ClBa_6$) octahedra surrounded by $(CO_3)^{2-}$ triangles, additional Ba^{2+} ions, and $[SiO_3]$ chains of corner-sharing SiO_4 tetrahedra (Figure 9). The high complexity ($I_{G,\text{total}} = 3913.746$ bits/u.c.) is created by the presence of both cation- and anion-centered[38] polyhedra

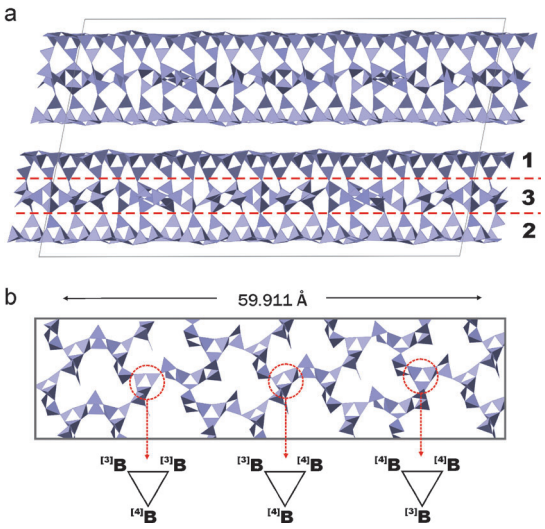

Figure 8. Arrangement of complex borate layers in the structure $Cs_3B_7O_{12}$ (a; each layer can be subdivided into three sublayers: **1**, **2**, and **3**) and projection of the layer **3** that shows that the groups indicated by red circles are translationally non-equivalent (b).

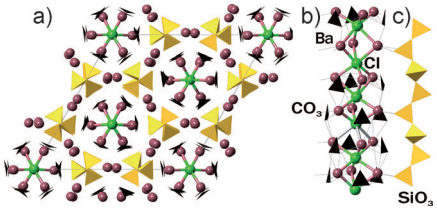

Figure 9. The crystal structure of $Ba_{12}(SiO_3)_6(CO_3)_5Cl_2$ (a) contains $[ClBa_3]^{5+}$ chains of face-sharing $ClBa_6$ octahedra surrounded by $(CO_3)_{2-}$ triangles (b), Ba^{2+} cations, and $[SiO_3]$ chains of corner-sharing SiO_4 tetrahedra (c).

as the basic building units and their mutual arrangement in the resulting structure.

In contrast to the oxo compounds, the number of very complex chalcogenides (e.g., sulfides and selenides) is much smaller. One of a few examples is the structure of ICF-26 (Figure 10a), consisting of two interpenetrating **mog**-type frameworks formed by corner-sharing of $[Li_4In_{22}S_{44}]^{18-}$ supertetrahedra (Figure 10b). Similar to zeolites, the high complexity ($I_{G,total} = 5444.699$ bits/u.c.) in this case is generated by the presence of nanoscale building blocks that self-assemble into a structure with relatively simple overall topology (Figure 10c).

5. Summary and Perspectives

Quantification of crystal-structure complexity using information theory allows basic mechanisms to be identified that are responsible for the appearance of complex structures in inorganic systems. The first and the most obvious one is a combination of chemically different components taken in complex proportions [e.g., in the intermetallic $Al_{55.4}Cu_{5.4}Ta_{39.1}$ (*ACT*-71) or $Ba_{12}(SiO_3)_6(CO_3)_5Cl_2$]. This approach usually

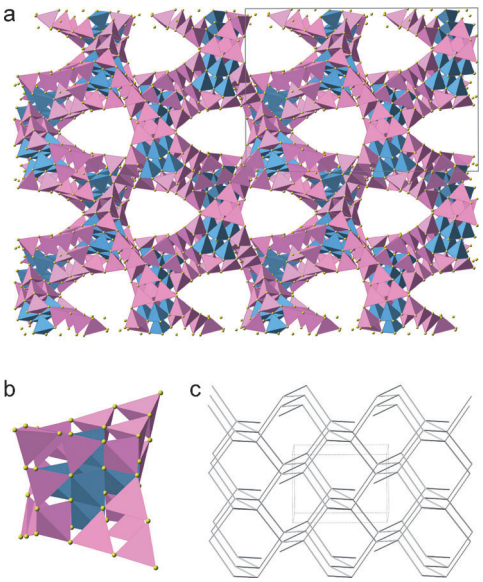

Figure 10. Structure of the In–Li sulfide ICF-26 composed from two interpenetrating frameworks (a) formed by corner-linkage of $[Li_4In_{22}S_{44}]^{18-}$ supertetrahedra (b). Both frameworks have a rather simple moganite topology (c). Pink InS_4 tetrahedra, blue LiS_4 tetrahedra.

leads to the association, in the same structure, of different building blocks that can be considered as being extracted from the parent solids of higher structural simplicity. This feature of inorganic structures is known as modularity and has been widely applied to the description of complex materials.[39] Another complexity-generating mechanism depends upon the presence in the crystallization media (solution, melt, gas, etc.) of clusters of high nuclearity that self-assemble to form highly ordered three-dimensional crystalline arrays. This group of complex structures includes different kinds of polyoxometallates, cluster compounds, and fullerene-containing inorganic solids. It is very probable that the same mechanism is at work in the crystallization of porous frameworks containing various types of large building blocks (e.g., zeolites and microporous sulfide frameworks). Finally, complexity may be generated by local rather than global bonding requirements in compounds with simple chemical compositions. In this case, topologically simple frameworks acquire high structural complexity owing to the local atomic ordering and formation of superstructures and superlattices on cooling.

Identification of structural complexity with information encoded in the atomic arrangement and its density allows relations between information and kinetics and thermodynamic parameters in crystalline materials to be quantitatively examined. For instance, quantitative verification can be made for the principle of "simplexity",[40] which states that, in the course of fast processes, such as spontaneous crystallization at non-equilibrium conditions, structurally simpler metastable phases form more easily that their more complex but stable counterparts.[16d] Another interesting problem is the behavior of information along the path of phase transitions induced by changing pressure and temperature. Probably, the most fundamental question of interest is the relation between information and energy[41] in processes that involve transformations of crystalline phases. It should be stressed that the

information content (or information entropies) of crystal structures considered herein are not identical to the thermodynamic entropies of the corresponding solids and to the information content employed in the statistical mechanics; the relations between these parameters are presently unclear. All these problems are far beyond the topic of this Minireview and require further exploration.

I am indebted to Prof. V. A. Blatov for incorporation of information-based complexity measures into the TOPOS program package. This work was supported by the Russian Federal Grant-in-Aid Program "Cadres" (agreement no. 8313) and Russian Foundation for Basic Research (grant 13-05-00928).

Received: May 21, 2013
Revised: July 6, 2013
Published online: December 13, 2013

[1] J.-M. Lehn, *Angew. Chem.* **2013**, *125*, 2906 – 2921; *Angew. Chem. Int. Ed.* **2013**, *52*, 2836 – 2850.
[2] a) J. Avery, *Information Theory and Evolution*, World Scientific, Singapore, **2003**; b) *Complexity in Chemistry, Biology, and Ecology* (Eds.: D. Bonchev, D. H. Rouvray), Springer, New York, **2005**; c) *Information and The Nature of Reality, From Physics and Metaphysics* (Eds.: P. Davies, N. Gregersen), Cambridge University Press, Cambridge, **2010**; d) D. J. Graham, *Chemical Thermodynamics and Information Theory with Applications*, CRC, Taylor & Francis, Boca Raton, **2011**.
[3] E. Shannon, W. Weaver, *The Mathematical Theory of Communications*, University of Illinois Press, Urbana, **1949**.
[4] a) O. Goldreich, *Computational Complexity: A Conceptual Perpective*, Cambridge University Press, Cambridge, **2008**; b) A. Sanjeev, B. Boaz, *Computational Complexity: A Modern Approach*, Cambridge University Press, Cambridge, **2009**.
[5] C. H. Bennet in *Complexity, Entropy, and the Physics of Information, Santa Fe Institute Studies in the Sciences of Complexity* (Ed.: W. H. Zurek), *Vol. VIII*, Addison-Wesley, **1990**, pp. 137 – 148.
[6] a) M. J. Buerger, V. Venkatakrishnan, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 4348 – 4351; b) S. E. Lister, I. R. Evans, J. A. K. Howard, A. Coelho, J. S. O. Evans, *Chem. Commun.* **2004**, 2540 – 2541; c) S. E. Lister, I. R. Evans, J. S. O. Evans, *Inorg. Chem.* **2009**, *48*, 9271 – 9281; d) L. Bindi, F. Nestola, A. Guastoni, L. Secco, *Mineral. Mag.* **2010**, *74*, 999 – 1012; e) C. Hoch, A. Simon, *Angew. Chem.* **2012**, *124*, 3316 – 3319; *Angew. Chem. Int. Ed.* **2012**, *51*, 3262 – 3265.
[7] J. K. Burdett, C. Marians, J. F. Mitchell, *Inorg. Chem.* **1994**, *33*, 1848 – 1856.
[8] L. Pauling, *J. Am. Chem. Soc.* **1929**, *51*, 1010 – 1026.
[9] F. C. Hawthorne in *Landmark Papers: Structure Topology* (Ed.: F. C. Hawthorne), The Mineralogical Society of Great Britain and Ireland, Twickenham, **2006**, p. L12.
[10] W. H. Baur, E. T. Tillmanns, W. Hofmeister, *Acta Crystallogr. Sect. B* **1983**, *39*, 669 – 674.
[11] A. L. Mackay, J. Klinowski, *Comp. Math. Appl. Part B* **1986**, *12*, 803 – 824.
[12] A. L. Mackay, *Crystallogr. Rep.* **2001**, *46*, 524 – 526.
[13] E. Estevez-Rams, R. González-Férez, *Z. Kristallogr.* **2009**, *224*, 179 – 184.
[14] W. Steurer, *Acta Crystallogr. Sect. A* **2011**, *A67*, C184.
[15] a) N. Rashevsky, *Bull. Math. Biophys.* **1955**, *17*, 229 – 235; b) E. Trucco, *Bull. Math. Biophys.* **1956**, *18*, 129 – 135; c) D. Bonchev, *SAR QSAR Environ. Res.* **2003**, *14*, 199 – 214; d) M. Dehmer, A. Mowshowitz, *Inform. Sci.* **2011**, *181*, 57 – 78.
[16] a) S. V. Krivovichev, *Acta Crystallogr. Sect. A* **2012**, *68*, 393 – 398; b) S. V. Krivovichev, *Struct. Chem.* **2012**, *23*, 1045 – 1052; c) S. V. Krivovichev, *Microporous Mesoporous Mater.* **2013**, *171*, 223 – 229; d) S. V. Krivovichev, *Miner. Mag.* **2013**, *77*, 275 – 326.
[17] T. J. R. Weakley, E. R. Ylvisaker, R. J. Yager, J. E. Stephens, R. D. Wiegel, M. Mengis, P. Wu, P. Photinos, S. C. Abrahams, *Acta Crystallogr. Sect. B* **2004**, *60*, 705 – 715.
[18] S. V. Krivovichev, E. V. Kir′yanova, S. K. Filatov, P. C. Burns, *Acta Crystallogr. Sect. C* **2000**, *56*, 629 – 630.
[19] G.-C. Guo, W.-D. Cheng, J.-T. Chen, H.-H. Zhuang, J.-S. Huang, Q.-E. Zhang, *Acta Crystallogr. Sect. C* **1995**, *51*, 2469 – 2471.
[20] A. K. Pant, *Acta Crystallogr. Sect. B* **1968**, *24*, 1077 – 1083.
[21] P. J. Mohr, B. N. Taylor, D. B. Newell, *Rev. Mod. Phys.* **2008**, *80*, 633 – 730.
[22] Inorganic Crystal Structure Database. FIZ Karlsruhe GmbH **2012**.
[23] V. A. Blatov, A. P. Shevchenko, V. N. Serezhkin, *J. Appl. Crystallogr.* **2000**, *33*, 1193.
[24] T. Weber, J. Dshemuchadse, M. Kobas, M. Conrad, B. Harbrecht, W. Steurer, *Acta Crystallogr. Sect. B* **2009**, *65*, 318 – 325.
[25] T. Mitra, P. Miró, A.-R. Tomsa, A. Merca, H. Bögge, J. B. Ávalos, J. M. Poblet, C. Bo, A. Müller, *Chem. Eur. J.* **2009**, *15*, 1844 – 1852.
[26] A. Müller, P. Gouzerh, *Chem. Soc. Rev.* **2012**, *41*, 7431 – 7463.
[27] a) P. C. Burns, *Miner. Mag.* **2011**, *75*, 1 – 25; b) J. Qiu, P. C. Burns, *Chem. Rev.* **2013**, *113*, 1097 – 1120.
[28] a) T. Z. Forbes, J. G. McAlpin, R. Murphy, P. C. Burns, *Angew. Chem.* **2008**, *120*, 2866 – 2869; *Angew. Chem. Int. Ed.* **2008**, *47*, 2824 – 2827; b) G. E. Sigmon, D. K. Unruh, J. Ling, B. Weaver, M. Ward, L. Pressprich, A. Simonetti, P. C. Burns, *Angew. Chem.* **2009**, *121*, 2775 – 2778; *Angew. Chem. Int. Ed.* **2009**, *48*, 2737 – 2740; c) J. Ling, C. M. Wallace, J. E. S. Szymanowski, P. C. Burns, *Angew. Chem.* **2010**, *122*, 7429 – 7431; *Angew. Chem. Int. Ed.* **2010**, *49*, 7271 – 7273.
[29] C. A. Ohlin, E. M. Villa, J. C. Fettinger, W. H. Casey, *Angew. Chem.* **2008**, *120*, 5716 – 5718; *Angew. Chem. Int. Ed.* **2008**, *47*, 5634 – 5636.
[30] B. S. Guiton, H. Wu, P. K. Davies, *Chem. Mater.* **2008**, *20*, 2860 – 2862.
[31] G. Costentin, A. Leclaire, M. M. Borel, A. Grandin, B. Raveau, *Z. Kristallogr.* **1992**, *201*, 53 – 58.
[32] E. K. Gordon, S. Samson, W. B. Kamb, *Science* **1966**, *154*, 1004 – 1007.
[33] A. Depla, E. Verheyen, A. Veyfeyken, M. Van Houteghem, K. Houthoofd, V. Van Speybroeck, M. Waroquier, C. E. A. Kirschhock, J. A. Martens, *J. Phys. Chem. C* **2011**, *115*, 11077 – 11088.
[34] C. Baerlocher, L. B. McCusker, D. H. Olson, *Atlas of Zeolite Framework Types, 6th ed.*, Elsevier, Amsterdam, **2007**.
[35] C. Baerlocher, T. Weber, L. B. McCusker, L. Palatinus, S. I. Zones, *Science* **2011**, *333*, 1134 – 1137.
[36] G. Nowogrocki, N. Penin, M. Touboul, *Solid State Sci.* **2003**, *5*, 795 – 803.
[37] A. M. Il′yinets, Yu. A. Malinovskii, *Dokl. Akad. Nauk SSSR* **1985**, *282*, 890 – 894.
[38] S. V. Krivovichev, O. Mentre, O. I. Siidra, M. Colmont, S. K. Filatov, *Chem. Rev.* **2013**, *113*, 6459 – 6535.
[39] a) G. Ferraris, E. Mackovicky, S. Merlino, *Crystallography of Modular Materials*, Oxford University Press, Oxford, **2004**; b) M. G. Kanatzidis, *Acc. Chem. Res.* **2005**, *38*, 359 – 368.
[40] J. R. Goldsmith, *J. Geol.* **1953**, *61*, 439 – 451.
[41] a) R. Landauer, *Phys. Lett. A* **1996**, *217*, 188 – 193; b) A. Bérut, A. Arakelyan, A. Petrosyan, S. Ciliberto, R. Dillenschneider, E. Lutz, *Nature* **2012**, *483*, 187 – 190.

