Article

Complexity Parameters for Molecular Solids

Alexander M. Banaru [1,2,*], Sergey M. Aksenov [2,3,*] and Sergey V. Krivovichev [4,5]

1. Faculty of Chemistry, Moscow State University, Vorobievy Hills, 119991 Moscow, Russia
2. Laboratory of Nature-Inspired Technologies and Environmental Safety of the Arctic, Kola Science Centre, Russian Academy of Sciences, 14 Fersman Street, 184209 Apatity, Russia
3. Geological Institute of Kola Science Centre, Russian Academy of Sciences, 14 Fersman Street, 184209 Apatity, Russia
4. Nanomaterials Research Centre of Kola Science Centre, Russian Academy of Sciences, 14 Fersman Street, 184209 Apatity, Russia; skrivovi@mail.ru
5. Department of Crystallography, Institute of Earth Sciences, Saint Petersburg State University, University Emb. 7/9, 199034 St. Petersburg, Russia
* Correspondence: banaru@mail.ru (A.M.B.); aks.crys@gmail.com (S.M.A.)

Abstract: Structural complexity measures based on Shannon information entropy are widely used for inorganic crystal structures. However, the application of these parameters for molecular crystals requires essential modification since atoms in inorganic compounds usually possess more degrees of freedom. In this work, a novel scheme for the calculation of complexity parameters (H_{molNet}, $H_{\text{molNet,tot}}$) for molecular crystals is proposed as a sum of the complexity of each molecule, the complexity of intermolecular contacts, and the combined complexity of both. This scheme is tested for several molecular crystal structures.

Keywords: information measure; complexity; crystal structure; crystallographic net; coordination number

1. Introduction

Nowadays, the most widespread method of assessment of the amount of information contained in a symbolic message is the calculation of so-called information entropy, introduced by Claude Shannon [1]. However, not only fragments of the text, but also other objects can be considered messages. The only condition necessary for the calculation of the information entropy falling on a "symbol" of some "message" is the knowledge of probabilities with which these "symbols" appear in the "message". The Shannon information theory approach was applied by Krivovichev for crystal structures, in which the role of a message is played by the list of the orbits (systems of symmetrically equivalent positions) occupied by atoms, while atoms themselves act as symbols. According to Krivovichev [2], the amount of structural information possessed by a single atom of a crystal structure equals the following:

$$I_{\text{G}}^{\text{str}} = -\sum_{i=1}^{k} p_i \log_2 p_i \ (\text{bits/atom}), \tag{1}$$

$$p_i = m_i/v, \tag{2}$$

where m_i is the multiplicity of the ith crystallographic orbit, v is the number of atoms in the reduced unit cell, and k is the number of crystallographic orbits in the crystal structure. That the cell should be strictly reduced as a repetition of the structural motif by translations actually does not add new information to that already contained. This indicator of complexity of a structure characterizes also the complexity of the quotient graph of the structure, i.e., the finite graph containing all v vertices of the reduced unit cell of the crystal structure along with the edges connecting the corresponding vertices in the crystal structure, and the edges connecting translationally equivalent vertices being turned into loops [3]. For instance, the quotient graphs of the crystal structures of α- and

β-quartz are identical (Figure 1). In both structures, Si and O atoms occupy one orbit each with the multiplicities equal to 3 and 6, respectively. Therefore, $I_G^{str} = -3/9 \cdot \log_2(3/9) -6/9 \cdot \log_2(6/9) \approx 0.918$ bits/atom.
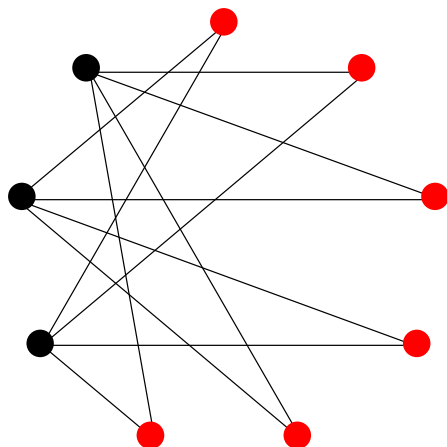

Figure 1. The quotient graph of α- and β-quartz. Black and red vertices correspond to the Si and O atoms, respectively.

The thermodynamic entropy of a crystal is the sum of configurational, vibrational, and other parts. The vibrational part of entropy has received significant attention, and several approaches have been developed for its quantitative estimation [4,5]. In general, aiming to obtain the vibrational entropy, high-quality multi-temperature high-resolution single-crystal X-ray diffraction analysis for the better refinement of anisotropic displacement parameters (ADPs) for all atoms, including hydrogens, which are known to be difficult to refine from X-ray diffraction data, together with complementary periodic DFT calculations, should be used [6]. However, it is worth recalling that ADPs are a dustbin for all experimental errors. Incorrect absorption correction, incomplete data, small amounts of disorder, diffuse scattering and thermal diffuse scattering, twining, problems with deconvolution of thermal motion, and charge density all can and will dramatically affect the ADPs. However, even assuming that the ADPs are perfect, it is not straightforward that the compound with larger ADPs will have the higher entropy, especially in the case of conformational polymorphs [6].

Considerably less attention has been paid to the evaluation of configurational entropy [7], except for the mixing entropy and order–disorder phase transitions [8,9]. Generally speaking, the configurational part of entropy represents a portion of its total entropy that originates solely from the location of its constituents and does not depend upon their dynamic properties. Within recent years, the vibrational spectra of strong and fragile metallic phases in different states (glass, liquid and crystalline) were directly measured in situ and reported [10]. For both strong and fragile phases, only a slight excess over the crystal was shown for the estimated vibrational entropies of liquid and glass, accounting for less than 5% of the total excess entropy estimated by step calorimetry. This result indicated that originally, the excess entropy of metals is almost totally configurational.

It was shown that the reduction in the configurational entropy of a crystal structure with respect to its greatest possible value is proportional to I_G^{str} [11], which can be considered a numerical confirmation of the obvious fact that making a crystal structure more complicated corresponds to a decrease in its configurational entropy.

The total information content of the reduced unit cell can be calculated as follows:

$$I_{G,\,tot}^{str} = -v \sum_{i=1}^{k} p_i \log_2 p_i \ (\text{bits/u.c.}), \tag{3}$$

The crystal structures of α- and β-quartz (Figure 1) both have $I_{G,\,tot}^{str} \approx 8.265$ bits/u.c. By the value of $I_{G,\,tot}^{str}$, crystal structures of inorganic substances can be classified as very

simple (<20 bits/u.c.), simple (20–100 bits/u.c.), intermediate (100–500 bits/u.c.), complex (500–1000 bits/u.c.), and very complex (>1000 bits/u.c.) [12]. The crystal structures of zeolites and microporous framework materials are commonly very complex. The most complex framework structure among minerals ever found is that of ewingite, $Mg_8Ca_8[(UO_2)_{24}(CO_3)_{30}O_4(OH)_{12}(H_2O)_8](H_2O)_{130}$ ($I_{G, tot}^{str}$ = 23,477.507 bits/u.c.) [13].

Borisov et al. [14] specified that a part of the unit cell that contains information on the crystal has the volume $V^* = V_0/M$, where V_0 is the volume of the unit cell, M is the point group order, and V^* is the volume of asymmetric unit (fundamental domain of the space group). In this case, the reduced unit cell should be considered, whereas the point group is understood as a crystal class. Within the asymmetric unit, some atoms can freely change their location (three degrees of freedom), while others can move only along a rotation axis or a mirror plane (one and two degrees of freedom, respectively) or have fixed positions (non-variant). According to [14], the ratio n_s/n_p, where n_s is the number of the "remained" translational degrees of freedom and n_p is the total number of atoms in the reduced unit cell, along with the value of V^*, characterizes the stability of a crystal structure. Along this line, it was proposed recently to complement the informational complexity of a crystal structure by its coordinational complexity [15], which considers the number of degrees of freedom of atoms, depending on their site-symmetry group. This measure also utilizes the classical entropy functional according to Shannon:

$$H = \sum_{i=1}^{k} L(p_i), \tag{4}$$

$$L(p_i) = \begin{cases} 0 \ (p_i = 0), \\ -p_i \log_2 p_i \ (p_i > 0) \end{cases}. \tag{5}$$

The coordinational complexity H_{coor} is calculated by formulas (4) and (5) with $p_i = a_i/A$, where $a_i \in \{0, 1, 2, 3\}$ is the number of degrees of freedom (arity) of the ith orbit, and $A = \sum_i a_i$. Analogous with H_{coor}, traditional information complexity I_G^{str} in the notations of [15] was rewritten as H_{comb} (combinatorial complexity), and the weighted sum of H_{comb} and H_{coor} (H_{conf}, configurational complexity):

$$H_{conf} = H(M, A) + \frac{M}{M + A} H_{comb} + \frac{A}{M + A} H_{coor}, \tag{6}$$

$$H(M, A) = -\frac{M}{M + A} \log_2 \frac{M}{M + A} - \frac{A}{M + A} \log_2 \frac{A}{M + A}, \tag{7}$$

$$H_{conf, tot} = (M + A) H_{conf}, \tag{8}$$

where $M = v$ (in Krivovichev's designations).

In fact, the value of H_{comb} characterizes the distribution of atoms over crystallographic orbits, H_{coor} is the distribution of translational degrees of freedom over occupied orbits, and H_{conf} deals with both distributions. Unlike the simple summation of information from two subsystems of any system, in expression (6), the property named strong additivity is used: aside from the terms accounting for two separate subsystems, this sum includes an additional term, $H(M, A)$, accounting for the informational gain followed from the mixing of the subsystems. The strong additivity is necessary in order for the informational index to fully satisfy the algebraic properties of the entropy functional [16]. Recently, for the calculation of H_{conf}, the software *crystIT* with the aid of Python language was created [17]. An important innovation of [17] was the option of calculating the complexity of a disordered structure with the partial population of orbits, which was not provided in any other scheme before.

The definition of the basic crystallographic terms used in this article can be found in the Online Dictionary of Crystallography compiled by IUCr [18]. From any point of the three-dimensional Euclidean space, the symmetry operations of a given space group *G* generate an infinite set of points, called a *crystallographic orbit*. The *site-symmetry group* (often called point symmetry) of a point is the finite group formed by the set of all symmetry

operations of G that leave that point invariant. It is isomorphic to a subgroup of the point group to which G belongs (the *crystal class*). A *Wyckoff position* of a space group G consists of all points for which the site-symmetry groups are conjugate subgroups of G. A *Wyckoff set* with respect to G is the set of all points for which the site-symmetry groups are conjugate subgroups of the normalizer N of G in the group of all affine mappings. The site-symmetry group of a Wyckoff position is the stabilizer of that position.

It is worth mentioning that each space group is characterized by the implicit informational complexity originated from Wyckoff positions [19]. The Wyckoff positions that can potentially be transformed into each other by the action of the normalizer of a space group form a Wyckoff set. This approach can be used for the calculation of p_i as the probability of some randomly chosen Wyckoff position to belong to the ith Wyckoff set. Similarly, any point group has its own implicit complexity following from the equivalence classes of group elements. With respect to the automorphisms of a group, some elements belong to the same class, and the partition of elements into classes may be referred to as the definition of p_i. The identity element of a group always forms its own class. For instance, in point group C_3, there is anidentity element and two rotations, $C_3{}^1$ and $C_3{}^{-1}$, forming another class, so the partition of C_3 is {1, 2}; then, $H = -1/3 \cdot \log_2(1/3) - 2/3 \cdot \log_2(2/3) \approx 0.918$ bits/element, and $H_{\text{tot}} = H \cdot 3 \approx 2.755$ bits/group. The 32 crystal classes are distributed over 18 abstract classes with a definite complexity (Table 1).

Table 1. Crystallographic point groups, abstract groups and their complexity following from the partition of group elements into equivalence classes with respect to automorphisms of the group.

Crystal Class (Schoenflies Symbols)	Crystal Class (Hermann–Mauguin Symbols)	Abstract Class	Order	Partition of Group Elements	H, Bits/Element	H_{tot}, Bits/Group
C_1	1	Z_1	1	{1}	0	0
C_2, C_i, C_s	$2, \bar{1}, m$	Z_2	2	{1, 1}	1.000	2.000
C_3	3	Z_3	3	{1, 2}	0.918	2.755
C_4, S_4	$4, \bar{4}$	Z_4	4	{1, 1, 2}	1.500	6.000
C_6, S_6, C_{3h}	$6, \bar{3}, \bar{6}$	Z_6	6	{1, 1, 2, 2}	1.918	11.510
C_{2h}, C_{2v}, D_2	$2/m, mm2, 222$	Dih_2	4	{1, 3}	0.811	3.245
D_{2h}	mmm	$Dih_2 \times Z_2$	8	{1, 7}	0.544	4.349
C_{3v}, D_3	$3m, 32$	Dih_3	6	{1, 2, 3}	1.459	8.755
C_{4h}	$4/m$	$Z_4 \times Z_2$	8	{1, 1, 2, 4}	1.750	14.000
C_{4v}, D_4, D_{2d}	$4mm, 422, \bar{4}2m$	Dih_4	8	{1, 1, 2, 4}	1.750	14.000
C_{6h}	$6/m$	$Z_4 \times Z_2$	12	{1, 2, 3, 6}	1.730	20.755
$C_{6v}, D_6, D_{3d}, D_{3h}$	$6mm, 622, \bar{3}m, \bar{6}m2$	Dih_6	12	{1, 1, 2, 2, 6}	1.959	23.510
D_{4h}	$4/mmm$	$Dih_4 \times Z_2$	16	{1, 1, 2, 4, 8}	1.875	30.000
D_{6h}	$6/mmm$	$Dih_6 \times Z_2$	24	{1, 2, 3, 6, 12}	1.865	44.755
T	23	A_4	12	{1, 3, 8}	1.189	14.265
T_h	$m\bar{3}$	$A_4 \times Z_2$	24	{1, 1, 6, 8, 8}	1.939	46.529
T_d, O	$\bar{4}3m, 432$	S_4	24	{1, 3, 6, 6, 8}	2.094	50.265
O_h	$m\bar{3}m$	$S_4 \times Z_2$	48	{1, 1, 3, 8, 8, 12, 15}	2.369	113.700

The value of H_{coor} and, as a result, the value of H_{conf}, is potentially capable of distinguishing crystal structures for which the values of H_{comb} are identical, for example, those of α- and β-quartz (Figure 1). In the crystal structure of α-quartz with the space group $P3_121$ (or $P3_221$), the Si atoms occupy general Wyckoff positions c with the stabilizer 1, and O atoms occupy the Wyckoff position a with the stabilizer 2. A general position is said to be tri-variant, as the atom can change its coordinates in three independent directions without loss of site symmetry, and the position on a twofold rotation axis 2 is mono-variant because, without the loss of its site-symmetry group, an atom can be displaced only along the direction of the axis. Therefore, $A = 3 + 1 = 4$. In the crystal structure of β-quartz with the space group $P6_222$ (or $P6_122$), the Si atoms occupy the Wyckoff position j (stabilizer 2), while the O atoms occupy the position c (stabilizer 222). The first position is mono-variant,

while the second is non-variant, from where it follows that $A = 1 + 0 = 1$. The values of H_{coor} in α- and β-quartz are, therefore, different (0.811 and 0 bits/atom).

　　If all atoms in the crystal structure occupy the same orbit (i.e., $k = 1$), then $H_{comb} = H_{coor} = 0$, and therefore, the two crystal structures containing one crystallographic orbit each cannot be distinguished from the viewpoint of their complexities. The examples are the simplest molecular crystal structures, such as iodine, rhombohedral sulfur S_6, α-N_2, which have a singlecrystallographic orbit each. It is worth noting that molecular crystals can be considered as having two independent sources of information: (i) the complexity of a molecule itself, and (ii) the complexity of molecular packing. The latter could be defined through the orbits formed by molecular mass centers. Very frequently, there is only one symmetrically independent molecule in a structure and therefore, again, $H_{comb} = H_{coor} = 0$. The value of H_{comb} calculated via the atomic orbits reflects the complexity of a molecule distorted by the intermolecular interactions but does not contain any information about the molecular packing. The value of H_{comb} calculated via the orbits of the molecular mass centers should reflect the complexity of molecular packing but is often equal to 0. An alternative way of conducting a complexity assessment for the molecular packing is to consider the orbits of the intermolecular contacts. An intermolecular contact is a pair of contacting molecules in a crystal [20], and its midpoint corresponds to a Wyckoff position with a certain stabilizer (site-symmetry group). In fact, H_{comb} calculated via the orbits of intermolecular contacts reflects the complexity of the edge net of a crystal. Combining the molecular complexity and the complexity of a packing, according to the strong additivity rule in the same manner as that given by (6), one would obtain a meaningful combinatorial complexity of a molecular crystal.

　　In this work, ten simple molecular crystal structures are considered in order to test the proposed model of complexity for molecular crystals. The selected structures include, ordered, I_2, S_6, α-S_8, α-N_2, β-P_4, C_{60}, ice Ih, polymorphic modifications I and II of benzene, and naphthalene (Table 2).

Table 2. Refcodes in ICSD/CSD/COD and crystallographic parameters for selected molecular crystals.

Refcode	Reference	Name	Space Group	Z''	Point Group of a Molecule	Wyckoff Sequence	Ideal Point Group of a Molecule
4511285	[21]	I_2	$Cmce$	1	$2/m$	f	∞/mmm
37090	[22]	S_6	$R\bar{3}$	1	$\bar{3}$	f	$\bar{3}m$
27840	[23]	α-S_8	$Fddd$	1	2	h^4	$\bar{8}2m$
15472	[24]	α-N_2	$Pa\bar{3}$	1	$\bar{3}$	c	∞/mmm
68326	[25]	β-P_4	$P\bar{1}$	3	1	i^{12}	$\bar{4}3m$
9011073	[26]	C_{60}	$Fm\bar{3}$	1	$m\bar{3}$	i^2h	$m\bar{3}5$
1011023	[27]	Ice Ih	$P6_3cm$	2	m	dc^4	$mm2$
BENZEN15	[28]	Benzene-I	$Pbca$	1	$\bar{1}$	c^6	$6/mmm$
BENZEN16	[28]	Benzene-II	$P2_1/c$	1	$\bar{1}$	e^6	$6/mmm$
NAPHTA11	[29]	Naphthalene	$P2_1/c$	1	$\bar{1}$	e^9	mmm

2. Materials and Methods

　　Crystallographic files *.cif were taken from the Inorganic Crystal Structure Database (ICSD) [30] and Crystallography Open Database (COD) [31] for inorganic molecular crystals, and the Cambridge Structural Database (CSD) [32] for organic molecular crystals. All computations were performed in the ToposPro software [33]. Net classifications were made on the basis of Reticular Chemistry Structure Resource (RCSR) [34].

3. Molecular Complexity

The approaches for measuring the information content of a molecule emerged in the early 1950s and were reviewed by Bonchev [35]. The most precise definition of vertex equivalence in a finite graph was first given by Trucco [36], and is based on the automorphism group of the graph. Equivalent graph vertices are those that belong to the same orbit of this group, and the derived index is called the information on the vertex orbits [37].

Let $p_i = n_i/N$, where N is the total number of atoms in a molecule, n_i is the multiplicity of the ith atomic orbit with respect to the point group of a molecule. Then, the functional (4) is called the molecular complexity H_{mol}, and

$$H_{\mathrm{mol,tot}} = N \cdot H_{\mathrm{mol}} \tag{9}$$

In fact, H_{mol} is equal to H_{comb}, but the former could be used not only for crystal structures, but also for separate molecules in a condensed or non-condensed phase (solution, gas, etc.). In the crystal structure of orthorhombic sulfur (α-S_8) with the space group *Fddd*, the stabilizer of the molecular mass center is 2 (Figure 2 left), while an idealized "crown" conformation has the symmetry $\bar{8}2m$ (Figure 2 right). In the crystal structure, atoms occupy four different orbits (Wyckoff positions h), $H_{\mathrm{mol}} = -4 \cdot 2/8 \cdot \log_2(2/8) = 2.000$ bits/atom, and $H_{\mathrm{mol,tot}} = H_{\mathrm{mol}} \cdot 8 \approx 16.000$ bits/molecule. Meanwhile, the idealized conformation has $H_{\mathrm{mol}} = 0$. Thus, a distortion of the molecule in the crystal accounts for 100% of the molecular complexity. It is noteworthy that the idealized symmetry is non-crystallographic, and therefore, the Wyckoff letters are not defined.
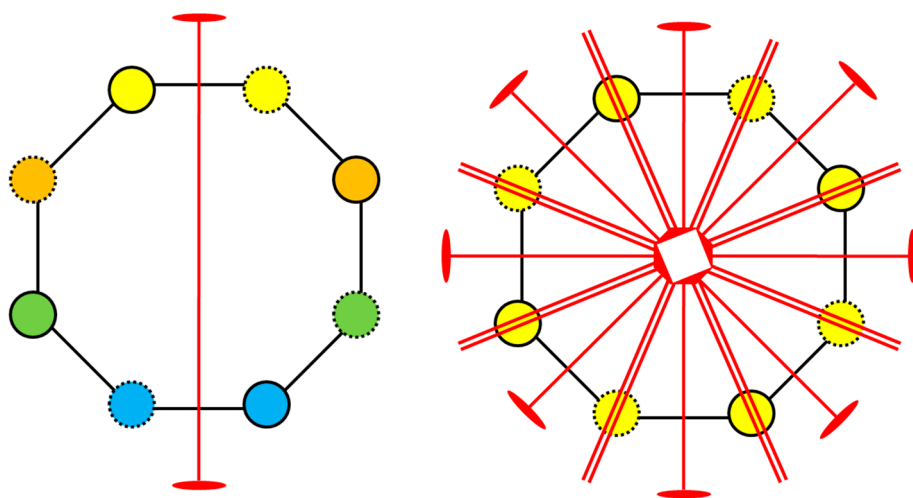

Figure 2. Symmetry elements of the S_8 ("crown") molecule in the structure of α-sulfur (**left**) and its highest symmetry (**right**). Atoms occupying different orbits are colored differently; atoms down the plane of the figure are shown by a dashed line.

For other selected structures, H_{mol} and $H_{\mathrm{mol,tot}}$ are listed in Table 3. Only 3 out of 10 structures (ice Ih, benzene, and naphthalene) have the positive H_{mol} values for their idealized symmetries, and 7 out of 10 for real symmetry in the crystal. Both benzene modifications (I and II) have the same stabilizer of the molecule, $\bar{1}$, and the same value of H_{mol}.

Table 3. The molecular complexity of selected molecular crystals H_{mol} (bits/atom) and $H_{mol,tot}$ (bits/molecule).

Name	Ideal Symmetry of a Molecule		Real Symmetry of a Molecule	
	H_{mol}	$H_{mol,tot}$	H_{mol}	$H_{mol,tot}$
I_2	0	0	0	0
S_6	0	0	0	0
α-S_8	0	0	2.000	16.000
α-N_2	0	0	0	0
β-P_4	0	0	3.585	14.340
C_{60}	0	0	1.522	91.320
Ice Ih	0.918	2.754	2.252	6.756
Benzene	1.000	12.000	2.585	31.020
Naphthalene	2.281	41.058	3.170	57.060

4. Edge Net Complexity

Graph edges can also be used as structural elements, and the orbits of the corresponding edge group of the graph can determine their equivalence classes as was first mentioned by Trucco [36]. This kind of information index is referred to as the information on the edge orbits of the graph [35]. To derive the edge net of a molecular crystal structures, it can be simplified in the following fashion:

1. At the first step, a molecular coordination shell is defined. In the ToposPro software, the coordination shell of a molecule is determined by means of molecular Voronoi–Dirichlet polyhedron (VDP) (for further details see [38]), but other ways of determination are also possible.
2. At the second step, each molecule is constricted to its mass center, and all mass centers are to be connected by edges with respect to the molecular coordination shell previously found. The number of edges adjacent to a node equals the molecular coordination number (CN_{mol}). If a crystal structure contains Z'' symmetrically not equivalent molecules, their CN_{mol} may differ, but in the selected crystal structures they do not.
3. At the third step, a midpoint is added to each edge of the net.
4. At the fourth step, two midpoints are to be connected if and only if they are adjacent to the same node; after that, the former nodes are to be deleted. The final net is called an edge net and is $2(CN_{mol} - 1)$-connected.

Figure 3 displays the edge net of ice Ih. Since the initial simplified net for Ih is 4-connected with the lonsdaleite topology(**lon**), the edge net occurs as being 6-connected with the extended lonsdaleite-type topology (**lon-e**). There are two symmetrically independent molecules in the crystal ($Z'' = 2$) and three symmetrically independent midpoints of intermolecular contacts (Figure 3 right). One of the occupied orbits corresponds to a Wyckoff position d (stabilizer 1, multiplicity 12), while the two other orbits correspond to c (stabilizer m, multiplicity 6) in the space group $P6_3cm$.

Let us call the edge net complexity (H_{edge}) a combinatorial complexity of the edge net. Then, $H_{edge} = -1/2 \cdot \log_2(1/2) - 2 \cdot 1/4 \cdot \log_2(1/4) \approx 1.500$ bits/contact. Each molecule forms CN_{mol} contacts with its neighbors, but each contact is shared by two molecules; therefore, the ratio of molecules and contacts is 1: $CN_{mol}/2$, and $H_{edge,tot} = 1.500 \cdot 4/2 = 3.000$ bits/molecule. Note that the symmetry of the edge net for Ih is not the maximal possible symmetry and thus, possesses an additional information content in comparison with the highest symmetry embedding of the net in 3D Euclidean space.
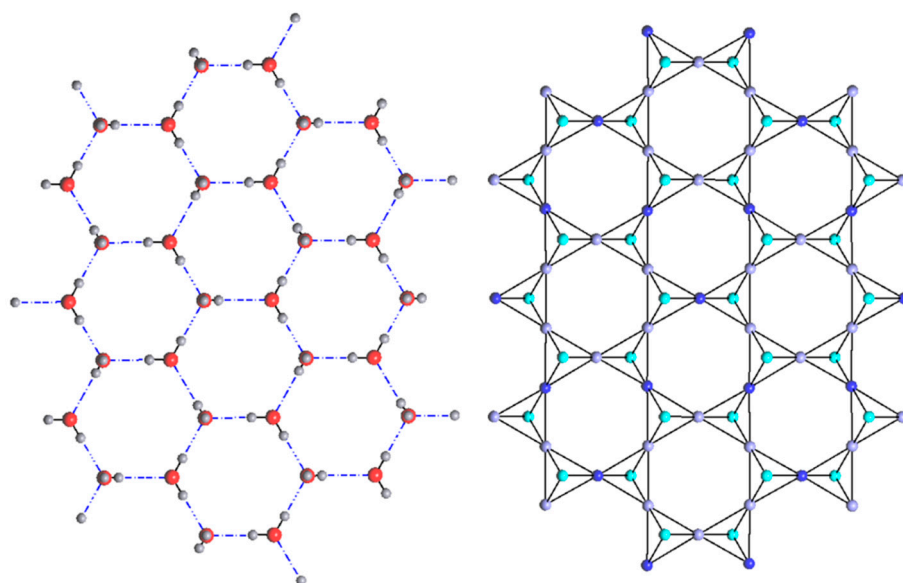
Figure 3. Crystal structure of ice (**left**) and H-bond midpoints (painted in shades of blue) in the edge net of the structure (**right**), view along [001]. Each shade corresponds to a certain crystallographic sort (an orbit) of midpoints.

An embedding of the highest symmetry may be generated by Systre software [39]. To find the most symmetric embedding of a net, the algorithm of the Systre software uses barycentric placement in which each vertex of the net is placed into the center of gravity of its adjacent vertices (all vertices have the same weight). The resulting net is called stable if no two vertices collide. Stable nets have maximal achievable crystallographic symmetry [40]. For the Ih edge net, the maximal possible symmetry is $P6_3/mmc$ (Table 4) as is that of lonsdaleite, with two orbits being occupied by edge midpoints: Wyckoff positions g (stabilizer $2/m$, multiplicity 6) and d (stabilizer $\bar{6}m2$, multiplicity 2). Then, $H_{edge} = -3/4 \cdot \log_2(3/4) - 1/4 \cdot \log_2(1/4) \approx 0.811$ bits/contact, $H_{edge,tot} \approx 1.622$ bits/molecule. This type of complexity is referred to as the *topological complexity* [11] of the edge net.

Table 4. Molecular coordination numbers (CN$_{mol}$) and the symmetry of edge nets for selected molecular crystals.

Name	CN$_{mol}$	Ideal Symmetry of Edge Net		Real Symmetry of Edge Net
		Space Group	**Wyckoff Sequence**	**Wyckoff Sequence**
I$_2$	14	$Im\bar{3}m$	cb	$fecb$
S$_6$	14	$Im\bar{3}m$	cb	eda
α-S$_8$	14 + 1	$Fddd$	$hgfedcb$	$hgfedca$
α-N$_2$	12	$Fm\bar{3}m$	d	d
β-P$_4$	14	$Fddd$	$fdcb$	$i^{17}hgfedcba$
C$_{60}$	12	$Fm\bar{3}m$	d	d
Ice Ih	4	$P6_3/mmc$	gd	dc^2
Benzene-I	12	$Fm\bar{3}m$	d	c^3
Benzene-II	12 + 2	$P4_2/mnm$	jcb	e^2dcb
Naphthalene	12 + 2	$P4_2/mnm$	jcb	e^2dcb

CN$_{mol}$ = 14 proved to be the most widespread among molecular crystals [41], which is the case for the majority of selected crystals (Table 4). The most frequent topology types are α-Fe or **bcu-x** (S$_6$, I$_2$), **14T3** (benzene-II, naphthalene), and **fcu** (α-N$_2$, C$_{60}$, benzene-II, as well as benzene-II and naphthalene, upon considering CN$_{mol}$ = 12). The ordered modification of white phosphorus (β-P$_4$) has another topology, **gpu-x**, which is more commonforrorganic crystals. The net of α-S$_8$ cannot be classified at CN$_{mol}$ = 15, but at CN$_{mol}$ = 14, it has the **14T191** topology according to the TopCryst database [42]. Never-

theless, all edge nets were generated at the highest found CN_{mol}. Note that α-S_8 has the most complex edge net among the selected structures (Table 5), being the only one with equal topological and real edge net positive complexities. Two other edge nets with equal topological and real edge net complexities (α-N_2 and C_{60}) have $H_{edge} = 0$. The largest contribution to a topological complexity occurs for the distorted β-P_4 (more than 170%), owing to the high value $Z'' = 3$.

Table 5. The edge net complexity of selected molecular crystals H_{edge} (bits/contact) and $H_{edge,tot}$ (bits/molecule).

Name	Ideal Symmetry of Edge Net		Real Symmetry of Edge Net	
	H_{edge}	$H_{edge,tot}$	H_{edge}	$H_{edge,tot}$
I_2	0.985	6.895	1.950	13.650
S_6	0.985	6.895	1.449	10.143
α-S_8	2.707	20.303	2.707	20.303
α-N_2	0	0	0	0
β-P_4	1.950	13.650	4.583	32.081
C_{60}	0	0	0	0
Ice Ih	0.811	1.622	1.500	3.000
Benzene-I	0	0	1.585	9.510
Benzene-II	1.379	9.653	2.236	15.652
Naphthalene	1.379	9.653	2.236	15.652

5. Molecular Net Complexity

Analogous with (6), H_{mol} and H_{edge} can be summed in a strong additive fashion. Since the molecular geometry and molecular packing are related, H_{mol} and H_{edge} are not purely independent, and their simple (non-weighted) sum should not be applied. The principal question for the use of the weighted sum is the values of the weights that should be attributed to H_{mol} and H_{edge}. Since they both may be calculated as an information content per molecule, a weighted scheme should be related to the ratio with respect to a molecule. In average, there are $CN_{mol}/2$ contacts per each N-atomic molecule, but, that the former number be ever an integer, the numbers $CN_{mol}/2$ and N should be multiplied by two:

$$H_{molNet} = H(2N, CN_{mol}) + \frac{2N}{2N + CN_{mol}} H_{mol} + \frac{CN_{mol}}{2N + CN_{mol}} H_{edge}, \tag{10}$$

$$H(2N, CN_{mol}) = -\frac{2N}{2N + CN_{mol}} \log_2 \frac{2N}{2N + CN_{mol}} - \frac{CN_{mol}}{2N + CN_{mol}} \log_2 \frac{CN_{mol}}{2N + CN_{mol}}, \tag{11}$$

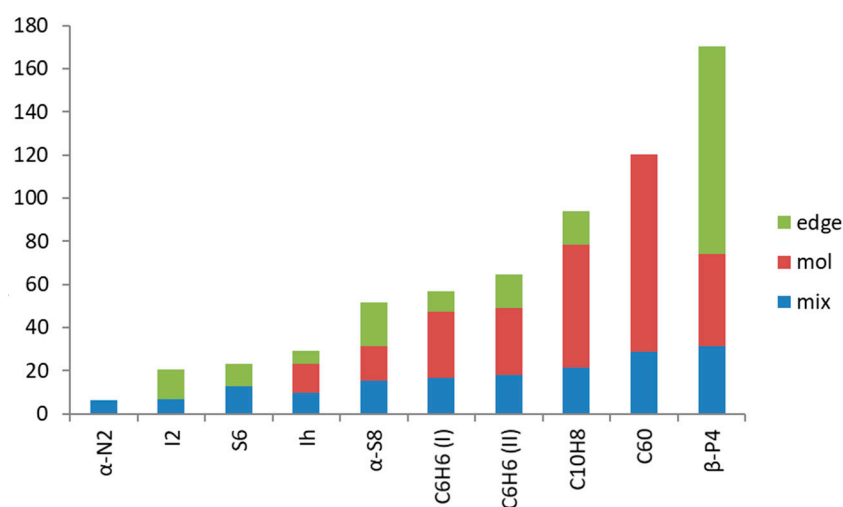$$H_{molNet, tot} = (N + CN_{mol}/2) H_{molNet}. \tag{12}$$

The values of H_{molNet} and $H_{molNet,tot}$ are expressed in bits per degree of freedom (bits/d.f.) and bits/molecule, respectively. A degree of freedom combines the ability of an atom to occupy a certain orbit with respect to the point group of the molecule and the ability of contacts to occupy a certain orbit with respect to the space group of the crystal. Each term in (10) reflects the impact of a specific source of complexity. The second and the third terms account for the impacts of H_{mol} and H_{edge}, respectively, while the first term accounts for the mixing of these two information sources. Note that the first term in (11) is determined by the ratio of N and CN_{mol} and is never 0, even when $H_{mol} = H_{edge} = 0$. Mixing two information sources is a source itself with meaningful complexity of the outcome.

Table 6 lists the molecular net complexities for the selected crystal structures. In case of $Z'' > 1$, $H_{molNet,tot}$ can be multiplied by Z'' in order to estimate a molecular complexity per Z'' molecule. As symmetrically independent molecules add to the total complexity independently, they act as distinct information sources. Contrastingly, a sole, symmetrically unique molecule replicated in the unit cell entails no additional information.

Table 6. The molecular net complexity of selected molecular crystals H_{molNet} (bits/d.f.) and $H_{\text{molNet,tot}}$ (bits/Z'' molecules).

Name	$H(2N, \text{CN}_{\text{mol}})$	H_{molNet}	$H_{\text{molNet,tot}}$
I_2	0.764	2.281	20.528
S_6	0.996	1.776	23.087
α-S_8	0.999	3.341	51.791
α-N_2	0.811	0.811	6.490
β-P_4	0.946	5.166	170.470
C_{60}	0.439	1.823	120.327
Ice Ih	0.971	2.922	29.222
Benzene-I	0.918	3.170	57.059
Benzene-II	0.949	3.406	64.712
Naphthalene	0.855	3.764	94.098

Among the selected structures, the highest value of $H(2N, \text{CN}_{\text{mol}}) \approx 0.999$ bits/d.f. is observed for α-S_8 (8 atoms, 7.5 contacts per molecule), and the lowest one is 0.439 bits/d.f. for C_{60} (60 atoms, 6 contacts per molecule). For the structure of α-N_2 $H(2N, \text{CN}_{\text{mol}}) \approx 0.811$ bits/d.f. is the only source of complexity (2 atoms, 6 contacts), and this structure has the lowest $H_{\text{molNet,tot}}$. The most complex structure is β-P_4. Two modifications of benzene essentially differ by the impact of $H_{\text{edge,tot}}$ and just slightly by the impact of $H(2N, \text{CN}_{\text{mol}})$ to a weighted sum (Figure 4).

Figure 4. $H_{\text{molNet,tot}}$ (bits/Z'' molecules) for the selected crystal structures in ascending order and separate weighted impacts of $H_{\text{edge,tot}}$ (edge), $H_{\text{mol,tot}}$ (mol) and $H(2N, \text{CN}_{\text{mol}})$ (mix) to $H_{\text{molNet,tot}}$.

In comparison with traditional combinatorial complexity, molecular net complexity has evidently more discriminating power. According to [43], the index of discrimination for typing methods, based on the probability of two unrelated objects being characterized as the same type, can be written as follows:

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{s} x_j (x_j - 1) \tag{13}$$

where N is the number of unrelated objects in the sample under consideration, s–the number of different types of objects in the sample, x_j–the number of objects belonging to the jth type, and D–the value of discriminatory power under assumption that objects are capable of being classified into mutually exclusive classes. For instance, $D = 1.0$ indicates a typing method being able to distinguish each object from all other objects of the sample. Similarly, $D = 0.0$ indicates all objects of the sample being of an identical type. If there is a

50% probability that the next object randomly chosen from the sample is indistinguishable from the first one, then $D = 0.5$. As one can see from Table 3, H_{mol} distinguishes just 7 of 10 selected structures ($D = 0.877$), whereas H_{molNet} distinguishes all of them ($D = 1.000$).

6. Conclusions

The calculation of complexity measures based upon Shannon information entropy has become a routine procedure for inorganic crystal structures [2,11,12,17]. However, a simple method of dividing atoms into equivalence classes with respect to their occupied orbits is of limited usefulness, especially for molecular crystals with a high symmetry of molecular constituents. Moreover, in contrast to common atomic crystal structures, molecular ones have two distinct sources of complexity: one resulting from a molecular geometry and the second one resulting from a molecular packing.

In this work, a common calculation scheme for the information of a crystal structure was adapted to molecular crystals by considering separate contributions of two complexity sources to the total complexity. As the sources are ruled by so-called strong additivity [15], a weighted scheme for the complexity with respect to one molecule was introduced, comprising molecular complexity, edge net complexity, and the combined complexity of the two sources. The scheme was tested for 10 selected molecular crystals and seems to be a powerful instrument for the assessment of complexity for molecular solids.

Author Contributions: Conceptualization, A.M.B.; formal analysis, A.M.B.; writing—original draft preparation, A.M.B. and S.M.A.; writing—review and editing, S.V.K.; supervision, S.V.K. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting reported results can be found in https://cloud.chem.msu.ru/s/cQ6YCRrb2MCQ6gM (accessed on 1 July 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Krivovichev, S.V. Topological Complexity of Crystal Structures: Quantitative Approach. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2012**, *68*, 393–398. [CrossRef] [PubMed]
3. Klee, W.E. Crystallographic Nets and Their Quotient Graphs. *Cryst. Res. Technol.* **2004**, *39*, 959–968. [CrossRef]
4. Fultz, B. Vibrational Thermodynamics of Materials. *Prog. Mater. Sci.* **2010**, *55*, 247–352. [CrossRef]
5. Benisek, A.; Dachs, E.; Salihović, M.; Paunovic, A.; Maier, M.E. The Vibrational and Configurational Entropy of α-Brass. *J. Chem. Thermodyn.* **2014**, *71*, 126–132. [CrossRef] [PubMed]
6. Kofoed, P.M.; Hoser, A.A.; Diness, F.; Capelli, S.C.; Madsen, A.Ø. X-Ray Diffraction Data as a Source of the Vibrational Free-Energy Contribution in Polymorphic Systems. *IUCrJ* **2019**, *6*, 558–571. [CrossRef]
7. Saito, K.; Yamamura, Y. Configurational Entropy and Possible Plateau Smaller than Rln2 in Complex Crystals. *Thermochim. Acta* **2005**, *431*, 21–23. [CrossRef]
8. Benisek, A.; Dachs, E. A Relationship to Estimate the Excess Entropy of Mixing: Application in Silicate Solid Solutions and Binary Alloys. *J. Alloys Compd.* **2012**, *527*, 127–131. [CrossRef]
9. Benisek, A.; Dachs, E. The Vibrational and Configurational Entropy of Disordering in Cu3Au. *J. Alloys Compd.* **2015**, *632*, 585–590. [CrossRef]
10. Smith, H.L.; Li, C.W.; Hoff, A.; Garrett, G.R.; Kim, D.S.; Yang, F.C.; Lucas, M.S.; Swan-Wood, T.; Lin, J.Y.Y.; Stone, M.B.; et al. Separating the Configurational and Vibrational Entropy Contributions in Metallic Glasses. *Nat. Phys.* **2017**, *13*, 900–905. [CrossRef]
11. Krivovichev, S.V. Structural Complexity and Configurational Entropy of Crystals. *Acta Crystallogr. Sect. B Struct. Sci.* **2016**, *72*, 274–276. [CrossRef]

12. Krivovichev, S.V. Which Inorganic Structures Are the Most Complex? *Angew. Chem. Int. Ed.* **2014**, *53*, 654–661. [CrossRef] [PubMed]

13. Krivovichev, S.V. Polyoxometalate Clusters in Minerals: Review and Complexity Analysis. *Acta Crystallogr.* **2020**, *76*, 618–629. [CrossRef] [PubMed]

14. Borisov, S.V.; Magarill, S.A.; Pervukhina, N.V. Crystallographic analysis of symmetry—stability relations in atomic structures. *J. Struct. Chem.* **2019**, *60*, 1191–1218. [CrossRef]

15. Hornfeck, W. On an Extension of Krivovichev ' s Complexity Measures. *Acta Crystallogr. Sect. A Found. Adv.* **2020**, *76*, 534–548. [CrossRef] [PubMed]

16. Csiszár, I. Axiomatic Characterizations of Information Measures. *Entropy* **2008**, *10*, 261–273. [CrossRef]

17. Kaußler, C.; Kieslich, G. CrystIT: Complexity and Configurational Entropy of Crystal Structures via Information Theory. *J. Appl. Crystallogr.* **2021**, *54*, 306–316. [CrossRef] [PubMed]

18. IUCr Online Dictionary of Crystallography. Available online: https://dictionary.iucr.org/Main_Page (accessed on 1 July 2021).

19. Banaru, A.M. Informational Entropy of Feodorov Groups. *Ural. Geol. J.* **2018**, *3*, 84–88.

20. Zorky, P.M. Symmetry, Pseudosymmetry and Hypersymmetry of Organic Crystals. *J. Mol. Struct.* **1996**, *374*, 9–28.

21. Bertolotti, F.; Shishkina, A.V.; Forni, A.; Gervasio, G.; Stash, A.I.; Tsirelson, V.G. Intermolecular Bonding Features in Solid Iodine. *Cryst. Growth Des.* **2014**, *14*, 3587–3595. [CrossRef]

22. Steidel, J.; Pickardt, J.; Steudel, R. Redetermination of the Crystal and Molecular Structure of Cyclohexasulfur, S6. *Z. Nature* **1978**, *33*, 1554–1555. [CrossRef]

23. Warren, B.E.; Burwell, J.T. The Structure of Rhombic Sulphur. *J. Chem. Phys.* **1935**, *3*, 6–8. [CrossRef]

24. Donohue, J. A Refinement of the Positional Parameter in α-Nitrogen. *Acta Crystallogr.* **1961**, *14*, 1000–1001. [CrossRef]

25. Simon, A.; Borrmann, H.; Craubner, H. Crystal Structure of Ordered White Phosphorus (β-P). *Phosphorus Sulfur Relat. Elem.* **1987**, *30*, 507–510. [CrossRef]

26. Dorset, D.L.; McCourt, M.P. Disorder and the Molecular Packing of C60 Buckminsterfullerene: A Direct Electron-Crystallographic Analysis. *Acta Crystallogr. Sect. A* **1994**, *50*, 344–351. [CrossRef]

27. Bernal, J.D.; Fowler, R.H. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *J. Chem. Phys.* **1933**, *1*, 515–548. [CrossRef]

28. Katrusiak, A.; Podsiadło, M.; Budzianowski, A. Association CH·π and No van Der Waals Contacts at the Lowest Limits of Crystalline Benzene I and II Stability Regions. *Cryst. Growth Des.* **2010**, *10*, 3461–3465. [CrossRef]

29. Cruickshank, D.W.J. A Detailed Refinement of the Crystal and Molecular Structure of Naphthalene. *Acta Crystallogr.* **1957**, *10*, 504–508. [CrossRef]

30. Hellenbrandt, M. The Inorganic Crystal Structure Database (ICSD)—Present and Future. *Crystallogr. Rev.* **2004**, *10*, 17–22. [CrossRef]

31. Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N.R.; Moeck, P.; Downs, R.T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420–D427. [CrossRef]

32. Groom, C.R.; Bruno, I.J.; Lightfoot, M.P.; Ward, S.C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179. [CrossRef] [PubMed]

33. Blatov, V.A.; Shevchenko, A.P.; Proserpio, D.M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Cryst. Growth Des.* **2014**, *14*, 3576–3586. [CrossRef]

34. O'Keeffe, M.; Peskov, M.A.; Ramsden, S.J.; Yaghi, O.M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* **2008**, *41*, 1782–1789. [CrossRef]

35. Bonchev, D. Shannon's information and complexity. In *Complexity: Introduction and Fundamentals*; CRC Press: London, UK, 2003; ISBN 978-0-429-16545-0.

36. Trucco, E. On the Information Content of Graphs: Compound Symbols; Different States for Each Point. *Bull. Math. Biophys.* **1956**, *18*, 237–253. [CrossRef]

37. Bonchev, D.; Mekenyan, O.V.; Trinajstić, N. Isomer Discrimination by Topological Information Approach. *J. Comput. Chem.* **1981**, *2*, 127–148. [CrossRef]

38. Blatov, V.A. Voronoi–Dirichlet Polyhedra in Crystal Chemistry: Theory and Applications. *Crystallogr. Rev.* **2004**, *10*, 249–318. [CrossRef]

39. Delgado-Friedrichs, O.; Hyde, S.T.; O'Keeffe, M.; Yaghi, O.M. Crystal Structures as Periodic Graphs: The Topological Genome and Graph Databases. *Struct. Chem.* **2017**, *28*, 39–44. [CrossRef]

40. Delgado-Friedrichs, O.; O'Keeffe, M. Identification of and Symmetry Computation for Crystal Nets. *Acta Crystallogr. Sect. A* **2003**, *59*, 351–360. [CrossRef] [PubMed]

41. Carugo, O.; Blatova, O.A.; Medrish, E.O.; Blatov, V.A.; Proserpio, D.M. Packing Topology in Crystals of Proteins and Small Molecules: A Comparison. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef] [PubMed]

42. The Samara Topological Data Center TopCryst. Available online: https://topcryst.com/ (accessed on 1 July 2021).

43. Hunter, P.R.; Gaston, M.A. Numerical Index of the Discriminatory Ability of Typing Systems: An Application of Simpson's Index of Diversity. *J. Clin. Microbiol.* **1988**, *26*, 2465–2466. [CrossRef]