

Classification of isentropic molecules in terms of Shannon entropy

Denis Sabirov^{1*} and Kamila Koledina¹

¹Laboratory of Mathematical Chemistry, Institute of Petrochemistry and Catalysis of RAS, 450075 Ufa, Russia

Abstract. Shannon entropy is widely used as a descriptor of molecular structure. Isentropicity of the molecules in terms of Shannon approach emerges in specific problems of physical and organic chemistry. In the present work, we consider the main classes of isentropic molecules, exemplify them, and provide their hierarchy.

1 Introduction

Shannon entropy (information entropy, h) is widely used as a structural descriptor in physical and organic chemistry. It allows assessing and comparing the complexities of the molecular structures [1–3]. For this purpose, a molecule is represented as disjoint subsets of atoms (atom types) made up with the chemically inequivalent atoms. The rules defining the inequivalence of the atoms in the molecule may vary but usually they are deduced from the analysis of the corresponding molecular graph [1–5]. In other words, the task of calculating Shannon entropy is based on counting the dissimilar atoms. The properties of this structural descriptor were deeply studied in previous works of the Bonchev and Basak groups (see their key papers [3,5]). In general, Shannon entropy provides numerical estimates of the molecules' complexity, reflects their symmetry, and correlates with important physical and chemical properties (*e.g.*, thermodynamic entropy or abundance) [2,3,5–9]. There are also works on the link between Shannon estimates and probabilities of chemical processes are discussed [9–13].

When molecules are treated in terms of the information-theoretic approach, the problem of comparison of may be emerged. Two marginal cases, when Shannon entropies are zero and maximal, were exhaustively discussed previously ($h = 0$ and $h_{\max} = \log_2 N$, where N is the total number of atoms in the molecule) [1–3, 9,10]. Working on the Shannon entropy of mixing molecules [1], we came to the conditions of equal entropies of the molecules and found that this issue has not been studied. In the present work, we treat, classify, and exemplify some cases isentropic molecules.

2 Computational details

The details of the calculations have been previously presented [1, 14]. Their main idea is based on the

partition of a molecule over N_1 atoms of the 1st type, N_2 atoms of the 2nd type, ... N_n atoms of the n -th type, where n is the number of atom types and $\sum N_j$ is the total number of atoms in the molecule. The Shannon entropy of the molecule in bits is introduced as a sum of the logarithms of the weights of each atom type ($N_j/\sum N_j$):

$$h = - \sum_{j=1}^n \frac{N_j}{\sum_{j=1}^n N_j} \log_2 \frac{N_j}{\sum_{j=1}^n N_j} \quad (1)$$

Two molecules A and B are isentropic if their information entropies are equal (N_A and N_B are the total numbers of atoms in the molecules):

$$\sum_{j=1}^{n_A} \frac{N_{Aj}}{N_A} \log_2 \frac{N_{Aj}}{N_A} = \sum_{i=1}^{n_B} \frac{N_{Bi}}{N_B} \log_2 \frac{N_{Bi}}{N_B} \quad (2)$$

Here, N_i , N_j , N_A , N_B , n , and m may be varied. Additionally, the partition of each molecule and the corresponding cardinalities of the subsets (N_i or N_j) are generally interdependent. Thus, we do not provide a general solution of Equation (2) and consider particular cases of the mentioned values when the condition of isentropicity is fulfilled.

3 Results and discussion

3.1 Zero-entropy molecules

If all atoms of the molecule belong to the single atom type, according to Equation (1), the Shannon entropy is zero. This is possible only under two simultaneous conditions:

- (a) A molecule is homonuclear (*i.e.* it is made up with the atoms of the one element).
- (b) The three-dimensional molecular structure reveals a high symmetry (so that all atoms are chemically equivalent).

* Corresponding author: diozno@mail.ru

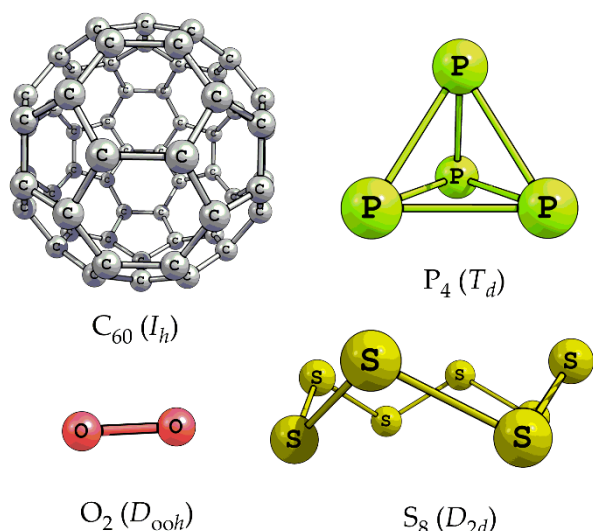


Fig. 1. Zero- h molecules. The point symmetry groups are shown in parentheses.

All zero- h molecules are isentropic and this is regardless of their molecular size. This size-independent class involves such different molecules become isentropic as O_2 , C_{60} , P_4 , and S_8 (Figure 1).

3.2 Isotomic molecules

Two molecules partitioned over the same number of atom types ($n_A = n_B$ in Equation (2)) we call *isotomic* (from ἴσος + τομή, ancient Greek ‘the same + partition’). Herewith, the molecular sizes (N_A and N_B) and cardinalities of the atom types (N_{A_j} and N_{B_i}) may be either equal or different. The condition of the isentropicity of isotomic molecules relates to the proportionality of the weights of all atom types in A and B:

$$\frac{N_{A_j}}{N_{B_i}} = \frac{N_A}{N_B} = k \quad (3)$$

We consider separately the cases when $k = 1$ and $k \neq 1$ (k is rational) in the subsections below.

3.2.1 Isotomic molecules with the same size

When $k = 1$, we deal with isotomic molecules having $N_A = N_B$. Trivially, this case corresponds to the molecules with identical molecular graphs. Therefore, we call them *isographic*. For example, CH_4 , CF_4 , CCl_4 , CBR_4 , and CI_4 are isentropic as the original information-theoretic approach puts that atoms of different chemical elements equally contribute to the entropy. Other examples of the same-size isentropic molecules with N up to 4 are provided below accompanying with the h values in bits and conventional designations of the molecule’s partition [number of atom types]×[number of atoms within them]:

- all heteronuclear diatomic particles AB (e.g., HF, HD, and HO*) – $h = 1$ (partition 2×1);
- particles with general formulas AAA (e.g., O_3 , N_3^- , and I_3^-) and ABA (e.g., H_2O , H_2S , and $:CH_2$) – $h = 0.918$ (partition $2 \times 1 + 1 \times 1$);

- ABC (e.g., HCN, HNC, and HOD) and AAB (e.g., N_2O) – $h = 1.585$ (partition 3×1);
- AB_3 (e.g., *CH_3 , NH_3 , PCl_3 , NO_3^- , and CO_3^{2-}) – $h = 0.811$ (partition $2 \times 1 + 1 \times 1$);
- ABBA (e.g., $HC \equiv CH$ and H_2O_2) – $h = 1$ (partition 2×2);
- A_2BC (e.g., $H_2C=O$ and $H_2C=S$) – $h = 1.5$ (partition $1 \times 2 + 2 \times 1$);
- ABBC, ABCD, and ABBB (e.g., $HC \equiv CCl$, $HCNO$, and $:C=C=C=O$, respectively) – $h = 2$ (partition 4×1).

Isentropic molecules of this class may be not isographic, *i.e.* have different molecular graphs but identical partitions (Figure 2). Consequently, within the simple information-theoretic approach, their Shannon entropies are equal. To discriminate such molecules numerically, the approach should be modified to take into account connectivity of the atoms (see, e.g. [15, 16]). To do or not to do this depends on the task of a study. It is noteworthy that the partitions obtained at the used level of the structural detailing correspond to NMR patterns of the molecules [17]. Usually, it is sufficient to solve typical problems of physical organic chemistry.

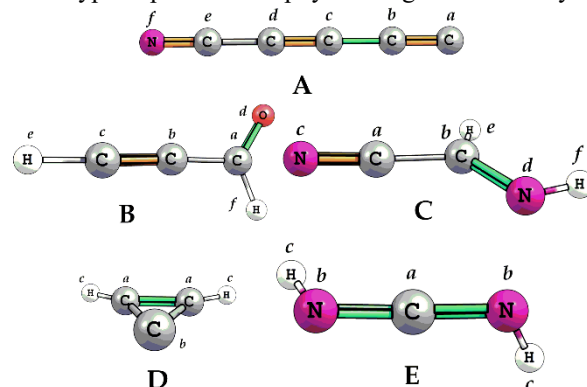


Fig. 2. Examples of isotomic molecules with the same size, partition pattern, and Shannon entropy but different molecular graphs. Molecules A–C consist of 6 atoms with partition 6×1 and $h = 2.585$ bits. Molecules D and E are penta-atomic with partition $2 \times 2 + 1 \times 1$ and $h = 1.522$ bits. Names of the molecules: cyanobuthadiynyl radical (A), propynal (B), *E*-cyanomethanimine (C), cyclopropenyliidene (D), and carbodiimide (E). Inequivalent atoms in the molecules are lettered.

3.2.2 Isotomic molecules with the different size

If k is rational and not unity, isotomic molecules reveal identical similarity in their molecular sizes (N_A and N_B) and the cardinalities of each atom types (N_{A_j} and N_{B_i}). This case have been discussed in [5]. We provide here two pairs of such *similar* molecules as:

- ozone O_3 (partition $1 \times 1 + 1 \times 2$) and ethylene C_2H_4 (partition $1 \times 2 + 1 \times 4$) with $k = 2$ and $h = 0.918$ bits;
- benzene C_6H_6 (partition 2×6) and acetylene C_2H_2 (partition 2×2) with $h = 1$ bit and $k = 3$.

Notably, similar molecules with equal Shannon entropies have pronouncedly different structure (Figure 3). At the same time, formation of the ensembles of such molecules reveals zero Shannon entropy of mixing even when the molecules differ.

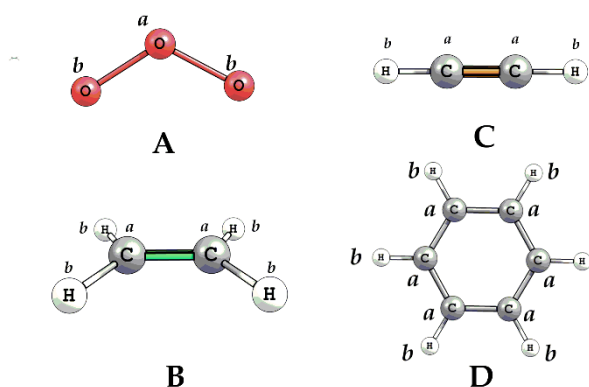


Fig. 3. Two pairs of similar isentropic molecules: ozone (A) and ethylene (B), acetylene (C) and benzene (D). Explanations are in the text. Inequivalent atoms in the molecules are lettered.

3.3 Allotomic molecules

We have occasionally found cases when two molecules with the same size and different partitions demonstrate equal Shannon entropies. For example, molecules with $N_A = N_B = 8$, partitions 4×2 and $1 \times 4 + 4 \times 1$ have $h_A = h_B = 2$ bits. The addition of another different atom to these molecular system leads to another pair of isentropic molecules: indeed, if $N_C = N_D = 9$, partitions $4 \times 2 + 1 \times 1$ and $1 \times 4 + 5 \times 1$, respectively, we have $h_A = h_B = 8/9 - 2 \log_2 3 \approx 2.281$ bits. The next addition also produces the isentropic pair. We try to exemplify these two cases with ‘real’ chemical systems (Figure 4). Even when we call such system real, these remain quite exotic (but existing). Such particles could be united under the name *allotomic* (from $\alpha\lambda\lambda\omicron\varsigma + \tau\omicron\mu\eta$, ancient Greek ‘different + partition’) and the search for extending examples of such chemical particles we address to our further studies.

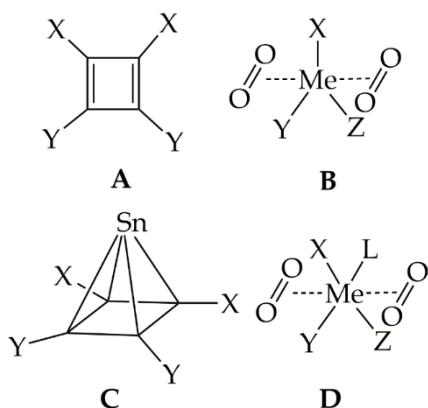


Fig. 4. Two pairs of allotomic isentropic particles. Explanations are in the text. Chemical nature of the particles: A is a derivative of cyclobutadiene, B and D are heteroligand metal complexes with molecular oxygen, C is an organotin compound with nonclassical arrangement of the atoms. Though compound C seems deliriously exotic, its analogues have been synthesized [18].

3.4 Tree of isentropic molecules and general remarks

To summarize, we provide a tree of isentropic molecules, which slightly differ from the plot of this

short paper (Figure 5). Indeed, one can see that the condition for zero Shannon entropy fits into Equation (2), so zero- h molecules are, in fact, isotomic and should not be allocated in a separate class.

We stress that this classification is valid within the original Shannon approach, in which the molecule is divided into atom types in line with the chemical equivalence of the atoms but the connectivity of the atoms does not directly influence on the calculated h values (but connectivity is used to discriminate the atoms over the atom types).

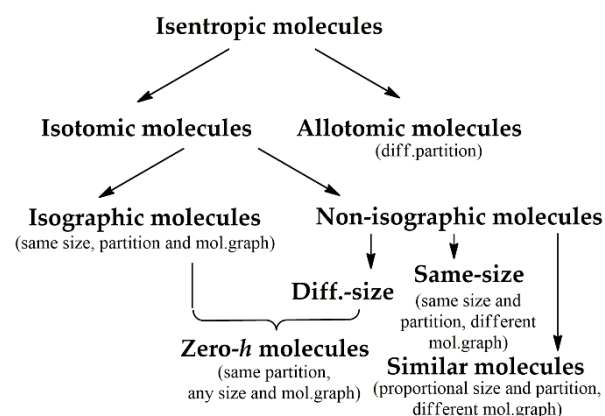


Fig. 5. Classification of isentropic molecules with the distinctive properties of their structures.

Isentropicity of the molecules within the information-theoretic approach should be further studied in the aspect of its physical meaning. As is known, Shannon approach has a disadvantage as it quantifies information but ignores its semantic properties [19]. Applied to the communication theory, it counts similar symbols in messages but ignores the sense, so two messages made up with the same set of the symbols have equal Shannon entropies but may differ in the sense. Something like that happens in the case of the application of Shannon approach to the molecular objects. In addition, there are questions when the above peculiarity is appropriate and when it is not and how Shannon isentropicity relates to physical/chemical properties of the molecules. Note that the approaches to quantifying semantic properties of physical objects within the information theory are developed [19].

4 Conclusions

In the present work, we have considered the main types of isentropic molecules and constructed the corresponding hierarchical tree to visualize their classification. We think that this result will be fruitful for quantification of the similarity/dissimilarity of molecular objects and further development of the information-theoretic structural descriptors.

The work has been performed under the theme ‘Novel theoretical approaches and software for modeling complex chemical processes and compounds with tunable physicochemical properties’ (registration number AAAA-A19-119022290011-6).

References

1. D. Sh. Sabirov, *Comput. Theor. Chem.* **1187**, 112933 (2020)
 2. D. Bonchev, O. V. Mekenyan, N. Trinajstić, J. *Comput. Chem.* **2**, 127 (1981)
 3. S. C. Basak, G. J. Niemi, G. D. Veith, *J. Math. Chem.* **7**, 243 (1991)
 4. S. J. Barigye, Y. Marrero-Ponce, F. Pérez-Giménez, D. Bonchev, *Mol. Divers.* **18**, 673 (2014)
 5. D. Bonchev, *Bulg. Chem. Commun.* **28**, 567 (1995)
 6. G. Castellano, J. L. González-Santander, A. Lara, F. Torrens, *Phytochemistry*. **93**, 182 (2013)
 7. M. Dehmer, K. Varmuza, S. Borgert, F. Emmert-Streib, *J. Chem. Inf. Model.* **49**, 1655 (2009)
 8. C. Raychaudhury, D. Pal, *Information Content Measures and Prediction of Physical Entropy of Organic Compounds* in book *Mathematical Foundations and Applications of Graph Entropy*, VI, eds M. Dehmer *et al.* (2017)
 9. D. Sh. Sabirov, E. Ōsawa, *J. Chem. Inf. Model.* **55**, 1576 (2015)
 10. D. Sh. Sabirov, *Comput. Theor. Chem.* **1097**, 83 (2016)
 11. D. Sh. Sabirov, A. O. Terentyev, V. I. Sokolov, *RSC Adv.* **6**, 72230 (2016)
 12. V. B. Aleskovskii, *Russ. J. Gen. Chem.* **72**, 569 (2002)
 13. V. M. Talanov, V. V. Ivanov, *Russ. J. Gen. Chem.* **83**, 2225 (2013)
 14. D. Sh. Sabirov, *Comput. Theor. Chem.* **1123**, 169 (2018)
 15. S. H. Bertz, *New J. Chem.* **27**, 860 (2003)
 16. T. Böttcher, *J. Chem. Inf. Model.* **56**, 462 (2016)
 17. D. Sh. Sabirov, O. Ori, I. László, *Fullerene Nanotube Carbon Nanostruct.* **26**, 100 (2018)
 18. V. Y. Lee, Y. Ito, A. Sekiguchi, H. Gornitzka, O. A. Gapurenko, V. I. Minkin, R. M. Minyaev, *J. Amer. Chem. Soc.* **155**, 8794 (2013)
- A. Kolchinsky, D. H. Wolpert, *Interface Focus* **8**, 20180041 (2018)