match no. 7 pp. 65 - 112 1979

INFORMATION INDICES FOR ATOMS AND MOLECULES

Danail Bonchev
Department of Physical Chemistry, Higher School of Chemical Technology, 8010 Burgas, Bulgaria

ABSTRACT

Information Theory is used as a basis for the introduction of new quantitative characteristics of atoms and molecules. In the case of the atoms of chemical elements this is performed according to the distribution of electrons of the electron shell over different subsets, determined by the values of quantum numbers. The nuclon distribution in the atomic nucleus according to their physical nature, spin and energy states, is separately considered. Two different approaches are applied to molecular species. Information indices are introduced for the distribution of atoms in the molecule according to their chemical nature, symmetry and topology. Other indices, characterizing the distribution of electron and atomic nuclei in the molecule, are independently introduced. The information quantities thus defined reflect fairly the structural features, and correlate with the basic properties of atoms and molecules.

INTRODUCTION

The notion of information appears as one of the most fundamental notions in the science of 20-th century, a notion of no less importance than that of matter and energy. This assertion follows from the very definition of information. According to Norbert Wiener[1] "information is neither matter, nor energy". Ashby[2] treats information as "a measure of the variety in a given system". Following Glushkov[3] "information is a measure of the non-homogeneity in distribution of matter or energy in space and time. On this basis it becomes more and more evident that besides their substantial and energetic essence, the objects and phenomena in nature and technics also have an information character. Moreover, the centre of scientific research is expected by some prognoses to move towards the information nature of the processes as the major field of study in 21-th century. These prognoses are essentially based on the possibility of systems and processes to be controlled by information, which is the major function of information in cybernetics. As a perspective these ideas could lead to a technology in which every atom or molecule is controlled by information, a possibility that is realized in living nature.

The year 1948 is usually considered as the birth date of Information Theory, since this was the year in which Claude Shannon published his fundamental work[4]. The concept of information as a quantity related to entropy is however much older. In 1894 Boltzmann[5] stated that every piece of information obtained for a physical system is related to the decrease in the number of its possible states, therefore the increase of entropy means "loss of information". In 1929 Szillard[6] developed this idea for the general case of information in physics. Later, Brillouin[7,8] generalised the concept of entropy and information in his negentropy principle of information. The possible interplay between Information Theory and Thermodynamics and in particular between entropy and information, is a subject of a permanent interest. (A selected bibliography in this field is given in reference 9). From the latest development of the problem one should especially mention the works of Kobozev[10] on thermodynamics of thinking, in which the concept of the anti-entropy character of thinking processes is developed.

Arising as "a special theory of communications", Information Theory soon exceeded its initial limits and found application in a number of scientific and technical areas: physics, chemistry, biology, medicine, linguistics, psychology, aesthetics et al. The role of information was recognized first in biology. Some important problems of conservation, processing and transmission of information in living beings were solved, such as coding of genetic information [11,12,13], estimation of the possibility of spontaneous selfgeneration of life on Earth[14], formulation of the fundamental laws of biological thermodynamics[15], analysis of the problems of bioenergetics[16] etc. The information content of systems was used as a quantitative criterion of evolution[17,18]. The information character of food-consuming processes was pointed out to be dominating over their substantial and energetic one[19,20]. Works on the information theory of disease diagnostics and therapyare in progress.

Information Theory is still of limited application in chemistry but the interest in this field is permanently growing. Levine, Bernstein et al.[21-24] developed an information-theoretic approach to molecular dynamics which describes the behaviour of interacting molecular systems far from equilibrium, in vibrationally and rotationally excited states. A quantitative measure of the information content, or entropy deficiency, of different classes of experiments is introduced, e.g. molecular beam scattering, IR chemiluminescence etc. This approach found various applications including the determination of branching ratios for alternative reaction paths, the study of the operational characteristics of lasers, etc. Daudel et al.[27-30] made use of the minimization of the information function upon partitioning of the space in molecular systems into a number of mutually exclusive spaces called loges, which contain localized groups of electrons. The sensitivity and catalytic activity of catalysts were connected with their information content by Kobozev et al.[31-34], who also formulated the optimum information conditions for characterization and prediction of catalyst properties. An information analysis was made of the surface treatment with different chemical reagents[35]. The formation and growth of crystals was treated as an information process [36,37]. This approach has a large application to geochemistry for the characterization of frequency distributions[38,39], the determination of complexity[40] and the general classifications of geochemical systems[41], etc. Information Theory was also applied to analytical chemistry for minimization of errors and analytical time, achievement of higher selectivity, estimation of the efficiency of analytical methods, etc.[42-47]. Gas chromatographic[48,49],atomic emission spectral analysis[50], and other analytical methods was handled in this way. The possibility of predicting the orientation in electrophylic atomatic substitutions was studied[51] on the basis of the comparison between information indices calculated from charge densities and the o, m, and p-topological orientation probabilities.

SOME REMARKS ON THE THEORETIC-INFORMATION APPROACH. THE CONCEPT OF STRUCTURAL INFORMATION

Information Theory offers quantitative methods for the study of receiving, conservation, processing, transmission and practical use of information. The quantitative measurement of information takes an important place in the information approach. The definition of the quantity of information requires a full rejection of the widespread but obscure ideas of information as the amount of facts, data, knowledge.

In the Shannon statistical information theory[4,7,52-55] the quantity of information is defined by means of the idea of probability. The probabilistic methods are of use in situations in which exists some uncertainty, about the choice of one or more elements from a certain set of them. The degree of uncertainty of a given outcome i is expressed by its entropy H_i as a function of the probability p_i of this result:

$$H_i = - \log p_i \qquad (1)$$

When the result is completely determined ($p_i = 1$), the entropy equals zero, whilst in the case of full uncertainty ($p_i = 0$) it is an infinite quantity (Indeed, $p_i \geq 0$, and $\sum p_i = 1$).

Another equation however is the principal one in the Shannon theory, that in which the mean entropy H(P) of the probability distribution $P = (p_1, p_2, \ldots, p_k)$ of all the possible outcomes (k in number) in a given situation is defined:

$$H(P) = - \sum_{i=1}^{k} p_i \log_2 p_i \qquad (2)$$

Here the logarithm at basis two is taken for measuring the entropy in bits (binary digits).

One should bear in mind that the notion of entropy as used in Information Theory, is more general than thermodynamis entropy. Viewed as a measure of the disorder in atomic and molecular motions, the thermodynamic entropy is a specific case of the general notion of entropy as a measure of any kind of disorder, or uncertainty, or uniformity.

Mostly, equation (2) is of use in the description of an experiment in which a given random variable takes k different values having probabilities $p_1, p_2, \ldots, p_k$. The quantity of information I is defined as the difference in the entropy values before and after the experiment:

$$I = H(P_0) - H(P_1) \qquad (3).$$

Thus, in the statistical information theory of Shannon, information is measured by means of the reduced uncertainty. When in an experiment the uncertainty is completely eliminated, the quantity of information equals the initial entropy:

$$I_{max} = H(P_0) = -\sum_{i=1}^{k} p_i \log_2 p_i \qquad (4).$$

If all P outcomes of the experiment are equiprobable, then

$$H(P) = \log_2 P \quad ; \quad I = \log_2 \frac{P_0}{P_1} \qquad (5),$$

where P_0 and P_1 denote the number of outcomes before and after the experiment. Again, in the case of a completely eliminated uncertainty ($P_1 = 1$), the information obtained is a maximum one, and is equal to the initial entropy:

$$I_{max} = \log_2 P_0 = H(P_0) \qquad (4').$$

Nonprobabilistic approaches to the quantitative definition of information are also possible. Ingarden and Urbanik [56,57] suggested such an axiomatic definition of Shannon's concept of information as a function of finite boolean rings. The so-called "epsilon-entropy" (an essentially combinatorial quantity), and especially the algorithmic quantity of information, both introduced by Kolmogorov [58-60], are of considerable interest. In the second case, the quantity of information is defined as a program of minimal length, allowing an one-to-one transformation of an object (set) into another one. The greater the difference between two objects, the larger the length of the program, and therefore the latter measures the degree of identity (or diversity) of these objects.

The nonprobabilistic methods in Information Theory extend the notion of quantity of information from a quantity of eliminated uncertainty to that of eliminated uniformity or to a quantity of variety, in agreement with the concept of Ashby[2]. It is in such a generalized interpretation that the information-theoretic approaches are applied to the determination of the so-called structural information - the quantity of information contained in the structure of a certain system.

Every structure is constructed of a certain number of elements (N) that could be, by means of a selected criterion of equivalency, distributed in subsets of equivalent elements. A finite probability scheme can be associated with this distribution:

$$\begin{pmatrix} N_1, & N_2, & \ldots, & N_k \\ p_1, & p_2, & \ldots, & p_k \end{pmatrix},$$

in which $p_i = N_i/N$ is the probability of a randomly chosen element to belong to the subset i having N_i elements. The ratio N_i/N can also be interpreted as the probability for a particular element of the structure to be involved in a given chemical reaction. Thus, the probability distribution as a whole can be viewed in terms of the freedom with which the structure will interact with other structures.

The entropy H(P) of the probability distribution of the elements of this structure, as defined by eqn.(2), can be viewed as a measure of the mean quantity of information, $\bar{I}$, contained in each element of the structure:

$$\bar{I} = -\sum_{i=1}^{k} p_i \log_2 p_i \text{ , bits/element} \qquad (6).$$

This statement needs to be discussed in detail. As pointed out by Rényi[61], H(P) from eqn.(2) can be interpreted either as a measure of entropy or as a measure of information. The first interpretation is justified when one deals with a system before an experiment is carried out on it. Thus, H(P) measures the uncertainty concerning the results of the experiment. When, however, one deals with a system after an experiment has been carried out on it, H(P) measures the amount of information obtained in the experiment. This is the case with various structures, including chemical ones, like atoms and molecules. The "experiment" reducing the uncertainty of the system is the very process of formation of a structure from the isolated elements. Here information is in bonded form, it is contained in the structure which is the reason to use the term "information content" of the structure.

Mowshowitz[62] considers additional arguments supporting the statement that the quantity $\bar{I}$, as defined for a certain structure from eqn.(6) is not a measure of entropy as the term is understood in Information Theory. It does not express the average uncertainty per structure having N elements, of a given ensemble of all possible structures having the same number of elements. Here, $\bar{I}$ <u>is rather the information content of the structure under consideration in relation to a system of transformations leaving the structure invariant</u>.

The specific character of the transformations, determining the kind of structural information, will be discussed in detail in the next chapters. Besides $\bar{I}$, that we shall further call a mean per element information content, the <u>total information content</u>[7] of the structure will be also of use:

$$I = N.\bar{I} = N \log_2 N - \sum_{i=1}^{k} N_i \log_2 N_i \text{ , bits} \qquad (7).$$

The concept of structural information based on the interpretation of eqns.(6) and (7), is consistent with the generalized view on the quantity of information as a quantity of variety[2]. There is no variety in a system composed of identical elements. In that case $k = 1$, $N_i = N$, and $I = \bar{I} = 0$ in eqns.(6) and (7), i.e. zero quantity of variety and zero quantity of information. In the other extremum case of a maximum variety of structural elements, $k = N$, $N_i = 1$, and the information content of the structure is a maximum one:

$$I_{max} = N \log_2 N \ , \quad \bar{I}_{max} = \log_2 N \qquad (8).$$

The quantities $\bar{I}$ and I from equs.(6) and (7) can be applied as structural indices expressing in a quantitative form the degree of complexity of the system. This makes them fairly suitable for diverse classification aims. They are a powerful tool for the discrimination between similar structures that may be of use in the coding of structural information. Another field of application of information indices is the correlation between structure and properties which could be successfully presented by information equations.

All this turns the information approach into an universal language for describing systems allowing useful analogies or common laws between systems of different nature to be inferred. In this way the laws of a given scientific area, properly translated into information language, could be projected into other, not well developed scientific areas (One could try, for instance, to find in biology analogies to the known chemical or physical laws). But the information approach is not simply a translation from less general languages to an universal one. It provides additional insight into systems and phenomena, allowing new results to be obtained, for instance on the basis of the extremum values of information indices. Various examples of the possible applications of this information approach will be given below.

ATOMIC INFORMATION INDICES

An atom represents a system whose structural elements: the protons, p; the neutrons, n; and electrons, z, are partitioned into two substructures: a nucleus having $n+p = A$ nuclons, and an electron shell of z electrons. According to equ.(7), a definite atomic information content[63] corresponds to this atomic structure:

$$I_{at} = (A+z) \log_2 (A+z) - z \log_2 z - A \log_2 A \quad (9)$$

In fact, equation (9) defines the atomic information carried by one of the isotopes of a given chemical element. Most of the chemical elements, however, represent a mixture of isotopes of diverse mass numbers A. Only about 20 elements are presented (with accuracy up to 0,001%) by one isotope. Due to this, the information content of a chemical element is defined[63] as an arithmetic mean of the atomic information of all their isotopes, $I_{at,\ i}$:

$$\bar{I}_{chem.elem.} = \sum_i I_{at,i} \cdot X_i \quad (10),$$

where X_i is the relative abundance of the isotope i in the chemical element.

The atomic information index, as defined by equs.(9) and (10), was calculated for the chemical elements from H to Pb ($z = 82$). It was applied to correlations with the properties of chemical elements[63], as well as an additional index in the definition of information content of molecular structures, especially those of heteroatomic organic compounds. However, a more complete definition of this index is necessary to extent its field of application. This can be done on the basis of the distribution of particles in the two atomic substructures - the nucleus and the electron shell. Thus, individual information indices could be introduced for atomic nucleus and electron shell:

$$I_{nucleus} = A \log_2 A - \sum_{i=1}^{k} p_i \log_2 p_i - \sum_{i=1}^{k'} n_i \log_2 n_i \qquad (11),$$

$$I_{electr.shell} = z \log_2 z - \sum_{i=1}^{k''} z_i \log_2 z_i \qquad (12).$$

The number of protons, neutrons and electrons, as well as the mass number, are denoted in equs. (11) and (12) by p, n, z, and A, respectively.

The <u>total atomic information content</u> is now defined as a sum of equs. (10), (11) and (12):

$$I_{at}^{total} = I_{at} + I_{nucl} + I_{el.shell} = (A+z) \log_2 (A+z) - \\ - \sum_{i=1}^{k} p_i \log_2 p_i - \sum_{i=1}^{k'} n_i \log_2 n_i - \sum_{i=1}^{k''} z_i \log_2 z_i \qquad (13).$$

The proton, neutron and electron distribution into different groups, k, k' and respectively k'' in number, is carried out by means of diverse criteria. For this reason, equations (11) and (12) will be further used for the definition of diverse information indices of atomic nucleus and electronic shell.

NUCLEAR INFORMATION INDICES

Following the nuclear shell model of Goeppert-Mayer and Jensen[64], the protons and neutrons in atomic nucleus can be viewed as being distributed on shells and subshells; according to the values of different quantum numbers: principal, angular momentum, spin, inner quantum number etc. An information index is associated with each of these distributions: I_{nucl}^{shell}, $I_{nucl}^{subshell}$, I_{nucl}^{n}, I_{nucl}^{l}, I_{nucl}^{s}, I_{nucl}^{j}, etc.

The features of the information indices thus defined were studied in detail[65]. They clearly reflect the regularities of the nucleus structure, and seem to be a good basis in the search of broader correlation between nuclear information indices and physical properties.

The importance of the information approach for the study of atomic nuclei can be also demonstrated by an additional nuclear index[66,67] - the <u>information on proton-neutron composition of nucleus</u>, I^{np}_{nucl}:

$$I^{np}_{nucl} = A \log_2 A - p \log_2 p - n \log_2 n \qquad (14).$$

This quantity is shown in fig. 1 as a function of the mass number for the 209 known 2β -stable isotopes. It alters

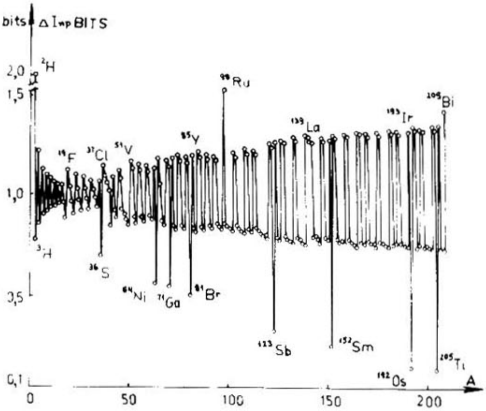


Fig.1. Dependence of the differential information on proton-neutron composition of the atomic nuclei on their mass number

harmoniously aproximating 1 bit, forming two branches with positive and respectively negative, deviations from the mean value 1 bit. The areas of different type of nucleus filling are set apart in fig.1: deutron type (n, p) for $A \leqslant 20$, and helium type (2n, 2p and 4n, 2p) at higher values of the mass number.

The total information for proton-neutron composition of nucleus is a magnitude quite near to the mass number A, remaining smaller than, or equal to, it. The difference in the two quantities is quite a small one:

$$I^{np'}_{nucl} = A - I^{np}_{nucl} = \frac{1}{2\ln 2} \cdot \frac{\beta^2}{A} = \frac{1}{2\ln 2} \frac{(A-2p)^2}{A} \qquad (15).$$

It can be called "defect" of the information on proton-neutron composition, since it expresses the loss of information upon atomic nucleus formation from free nuclons, and recalls the defect of the mass upon the same process. The two defects could be expected to be connected, since they both increase at higher deviations of the atomic nucleus from the symmetrical state with an equal number of protons and neutrons. This idea is supported by the coincidence (with an accuracy to a coefficient) between the defect of information (equ.15) and the so called parameter of the relative symmetry of the nucleus, δ, in Weizsäcker's equation[68] for the energy of atomic nucleus:

$$\delta = - \varepsilon \frac{(A-2p)^2}{A}, \quad \text{energy} \qquad (16).$$

In addition, correlation were found for isodifferent groups of nuclei (β = n-p = const) between binding energy and the total information on proton-neutron composition of atomic nuclei:

$$E_b = a \cdot I^{np}_{nucl(id)} + b \qquad (17).$$

Here a and b are constants for a given isodifferent group.

The information indices for proton-neutron composition of atomic nucleus was also applied to nuclide systematics, introdicing a new fifth kind of nuclides besides the known four kinds of nuclides[69]: isotopes ($p = const$); isotones ($n = const$); isobars ($A = const$); and isodifferentes ($\beta = const$). A theoretic-information variant of nuclide systematics was suggested[70] on the basis of the mean per nuclon defect of information on proton-neutron nuclear composition, $i = \triangle I^{np'}_{nucl}/A$. It represents (fig.2) a diagram of the mean defect vs. mass number (i/A-diagram) in which each of the nuclei is placed on a crossing of the five lines. The new type of nuclide group includes the nuclei of the chemical elements having the same mean defect of information. As a matter of fact these are new lines of nuclear genesis within which every nucleus can be obtained by a reaction of nuclear fusion between the initial nuclei.

The complete i/A-diagram, including all known 1723 isotopes of chemical elements[71], has a specific trend expressing the features of either nuclear and electron structure of the atom. The analogy in the nuclear and electron properties of chemical elements, substantiated in detail by means of the information indices, also allowed a good correlation to be found[70] between the nuclear binding energy[71] and electron energy[72] of isodifferent groups of nuclides.

One can conclude that nuclear information indices proved to be of use in the description of nuclear structure and in the systematics of atomic nuclei, for some correlations with the nuclear properties, as well as in the establishing of the information analogy in the periodic structure of an atomic nucleus and an electron shell.

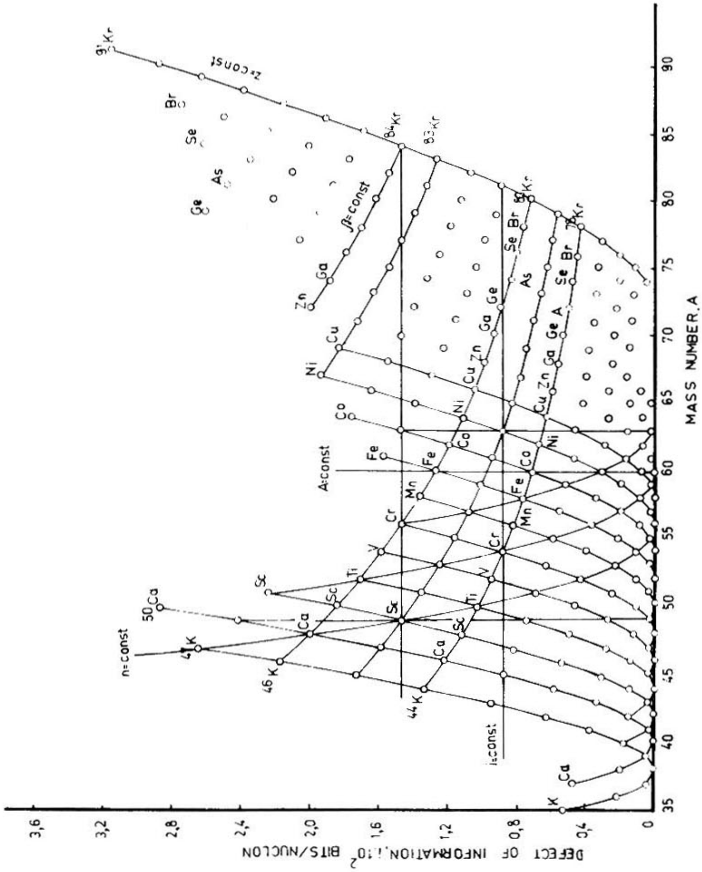


Fig.2. Dependence of the defect of information per nuclon on the mass number of nuclides of Period 4

INFORMATION INDICES FOR ELECTRON SHELLS OF ATOMS

The electrons in the atomic electron shell could be treated, like the nuclons in the atomic nuclei, as being distributed in different subsets: z $(z_1, z_2, \ldots, z_k)$. On this basis the mean per electron, as well as the total <u>information on electron distribution in the atomic electron shell</u> is introduced[73,74]:

$$\bar{I}^k_{\mathrm{el.shell}} = -\sum_{i=1}^{k} \frac{z_i}{z} \log_2 \frac{z_i}{z}$$
$$I^k_{\mathrm{el.shell}} = z \log_2 z - \sum_{i=1}^{k} z_i \log_2 z_i \qquad (18)$$

The kind of information index, k, is determined by the criterion selected for electron partitioning in subsets. Within one-electron approximation such criteria could be selected from the diverse atomic quantum numbers, as well as from some of their combinations determining different electronic groups or atomic energy states. The following types of electron subsets in the atom have been taken into consideration up to now[59,60]: 1. electron shells; 2. subshells; 3. atomic orbitals; 4. spin-orbitals; 5. nlj-subshells, as well as the groups of electrons having the same value of a given quantum number; 6. principal (n); 7. angular momentum (l); 8. magnetic (m); 9. magnetic spin (m_s); 10. inner (j); 11. total magnetic (m_j) and 12. the sum of the principal and angular momentum quantum numbers (n+l).

Consider, as an illustration of the information approach, the element chlorine having electron configuration $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^5$. The 17 electrons are distributed in the following subsets: shells (2, 8, 7), subshells (2, 2, 6, 2, 5), nlj-subshells (2, 2, 2, 4, 2, 2, 3), atomic orbitals (8x2, 1), spin-orbitals (17x1); by the values of the quantum numbers: (6, 11) for l = 0, 1; (10, 4, 3) for m = 0, ±1; (9, 8) for

$m_s = \pm 1/2$; (10, 7) for $j = 1/2, 3/2$; (2, 7, 7, 1) for $m_j = -3/2, -1/2, +1/2, +3/2$; (2, 2, 8, 5) for $n+l = 1, 2, 3, 4$. An information index, determined by equ.(18), corresponds to each of these 12 distributions.

Comparing the first five kinds of atomic information indices one could obtain the following inequalities (for simplicity the subscript for electron shell will be omitted):

$$I_n < I_{nl} < I_{nlj} < I_{nlm} < I_{nlmm_s} \qquad (19).$$

Inequalities (19) show that the more complete definition of electron energy by means of a subsequent introduction of additional quantum numbers decreases the uncertainty in energy, and therefore increases the information content of the atom. When electrons are treated as being distributed over spin-orbitals any uncertainty is eliminated and the quantity of information is a maximum one:

$$I_{nlmm_s} = N \log_2 N = I_{max} \qquad (20).$$

The information on electron distribution over spin (or otherwise spin-information), I_{m_s}, manifests interesting features. For all the elements with a closed electron shell it is exactly equal to the atomic number of the element, i.e. every electron in these elements carries 1 bit spin-information:

$$I_{m_s}^{\uparrow\downarrow} = z \text{ bits/atom} \; ; \; \bar{I}_{m_s}^{\uparrow\downarrow} = 1 \text{ bit/electron} \qquad (21).$$

The <u>differential information indices</u>:

$$\triangle I_k = I_{k,z} - I_{k,z-1} \qquad (22),$$

giving the increase in the information content of an atom of a chemical element related to an atom of another element, whose atomic number is smaller by one, are more sensitive atomic indices than those discussed above. As can be shown these functions have maximum in every element, in which the formation of a new electron subset begins. They decrease regularly with the increase in the number of electrons in this subset and have a minimum when the subset is maximally filled with electrons. In this way the differential information indices seem to be a convenient means for the description of the periodicy in the electronic structure of atoms.

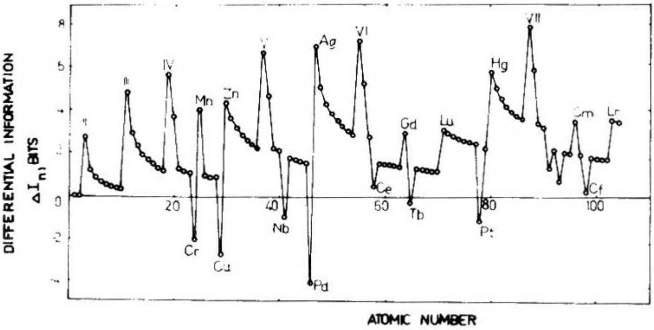


Fig.3. Differential information on electron distribution over shells in the atoms of chemical elements

This statement can be exemplified in fig.3 where the differential information on electron distribution over shells is shown for the chemical elements up to $z = 103$. This index has a maximum in the initial element of every period, since it is in this element that the filling of a new electron shell begins. A curve, common for a given period and ending in a minimum in the corresponding noble gas, is obtained for the second and third period only. In the other cases the information curve of every period is divided into different s, p, d, and f-parts, due to the delay in filling f- and d-subshells.

As the type of valence electrons determines the grouping of chemical elements into main, secondary and tertiary (lanthanides, actinides) groups, one can also conclude that the information index $\triangle I_n$ is capable of expressing correctly the horizontal and vertical structure of the Periodic Table. In fact, the elements of every main or secondary group can be connected in fig.3 in a common curve, to which a group information equation corresponds. Such equations were also deduced[75,76] on the basis of the differential indices on electron distribution over subshells, atomic orbitals, and the values of angular momentum quantum number. In addition, an equation was derived on the basis of the spin information that describes the groups and periods in the Periodic Table simultaneously:

$$\bar{I}_s = 1 - \frac{a^2}{2\ln 2(z_0+n)^2} \qquad (23),$$

where z_0 is a constant for every period, equal to the atomic number of the element concluding the preceding period, $n = z-z_0$ can be handled as the atomic number of the element in the period, whilst the number of single electrons, a, is a constant for a given group of chemical elements.

<u>Information indices</u> can be directly specified <u>for the periods and s, p, d, and f-subperiods</u> in the Periodic Table of chemical elements[67] as the total increase of atomic information content in these groupings of elements:

$$I_k^{\text{period}} = \sum_{\text{period}} \triangle I_{k,i} \; ; \; I_k^{\text{subperiod}} = \sum_{\text{subperiod}} \triangle I_{k,i} \qquad (24),$$

where $\triangle I_{k,i}$ is the differential index of the type k for the element i of the period (subperiod) under consideration, as defined by equ.(22).

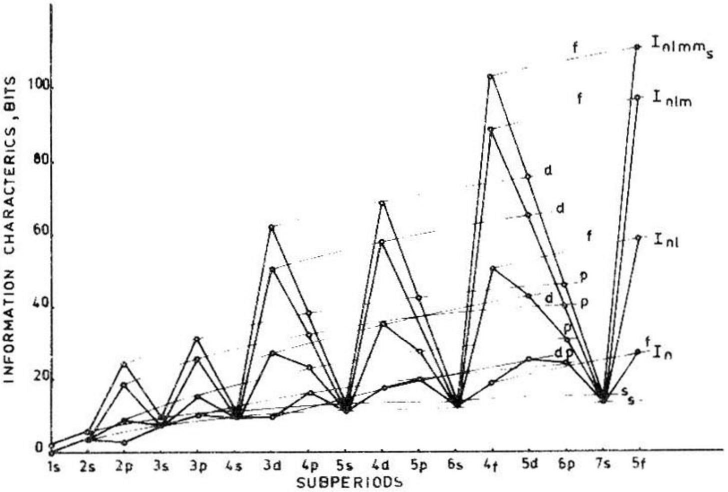


Fig.4. Four information indices for the subperiods in the Periodic Table of chemical elements

The information indices of all s, p, d, and f-subperiods of the Periodic Table are shown in fig.4. All subperiods of a given l-type are represented by common approximately paralel lines, the greater l is, the higher these lines lie. The regularity in the arrangment of periods in assending characteristic "waves" each one including a pair of periods, is a impressive one. It is worthwhile mentioning that the electron subshells are situated on the right side of the maxima in fig.4 in the sequence $(n-2)f \rightarrow (n-1)d \rightarrow np \rightarrow (n+1)s$, i.e. following the rules of Klechkovski[78] for electron subshells filling within the so-called (n+1)-electron groups.

The information indices of 4f, 5f, and 5d-subperiods, as well as the differential information on electron distribution over shells, were used in the analysis of the controversial problem of the exact position of lanthanides and actinides in the Periodic Table. They provided evidences in support of the Villar concept[79] according to which lanthanides include the elements from La to Yb (No 57 - 70) instead of Ce to Lu (No 58 - 71), and by analogy the actinides - from Ac to No (No 89 - 102) instead of (No 90 - 103).

The atomic information indices allow an opinion to be given to another controversial problem in the theory of Periodic Table - in which element the g-electron will first appear. A prediction was made[73], on the basis of the extrapolation of the information index for electron distribution on the values of angular momentum quantum number, that this could occur in the element having atomic number $z = 122$ or, less probably, $z = 123$.

Some other results were found for the excited states of atoms[80]. As a rule the light emission is accompanied by a decrease, while the light absorption - by an increase in the atomic information content:

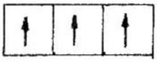
$$\text{Emission: } \Delta I \leqslant 0 \text{ ; Absorption: } \Delta I \geqslant 0 \qquad (25),$$

where the equality holds only for electron transitions from mono occupied to unoccopied energy levels. Proceeding from the concept of information as a negative component of entropy, as well as taking into account equ.(25), one could conclude that emission is accompanied by an increase in entropy, and therefore it is an irreversible process. This conclusion is consistent with the results of Andrade e Silva and Lochak[81], obtained by means of thermodynamics of isolated particle of L. de Broglie[82].

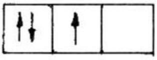
The information analysis of the states with inverse electron population provides another result of interest: The pairs of atomic (as well as molecular) states with equal positive and negative absolute temperature are characterized by equal information content.

The information indices, introduced for the electron shells of chemical elements, can be applied to various correlations with the properties of chemical elements. As it will be demonstrated in another report of this conference[83], for each of the properties of the elements of a certain group in the Periodic Table an information index exists correlating better to it than does the atomic number of elements. On this basis more accurate predictions can be made for the properties of the hypothetical superactinides of the end of period VII, and the beginning of period VIII.

Some generalizations are also possible. Thus the first Hund rule, requiring an initial population with one electron in every degenerate atomic (as well as molecular) orbital, appears to be associated with the requirement for a maximum information on the electron distribution over the atomic (or molecular) orbitals[84]. For instance:



$I_{AO(MO)} = \max$



$I_{AO(MO)} < \max$

The Pauli exclusion principle, forbidding a spin-orbital to be populated by more than one electron, can be interpreted in a similar manner[84]. As has already been shown by equ.(20), a maximum information on electron distribution over spin orbitals follows from this restriction. Thus, in accordance with the Ashby concept of information[2], the Pauli exclusion principle appears as a principle for aquiring maximum variety in atoms and molecules. This interpretation can be additionally supported by an information analysis of electronic wave functions, as well as of the irreducible representations of the symmetrical group S_N.

One could suppose that the Pauli exclusion principle and the first Hund rule are specific cases of a certain more general principle of maximum information content of matter. Other known or unknown rules or laws might also be associated with this general principle.

Concluding this chapter an opinion can be expressed that as a whole the concept of information content of chemical elements is a convenient mathematical model which reflects the real properties of the chemical elements and is likely to be homomorphic to the Periodic Table. No other mathematical function could manifest in such a drastic way the difference in the properties of chemical elements, as well as their periodicity. The information approach could be of use in the analysis of quantitative aspects of periodicity, in the search for correlations with the properties of chemical elements, as a mean of solving controversial problems in the theory of atomic structure and the Periodic Table etc. The formulated principle of maximum information content could provide the finding of new regularities in chemistry and other scientific areas.

INFORMATION INDICES FOR MOLECULES

The application of Information Theory to molecular objects is dated from the fifties when an information analysis of some biologically important molecules, as well as of their role in the living organisms were first made. In what follows we shall consider the known methods for the determination of the molecular information content subsequently examining the diverse criteria for equivalency of the atoms in a molecule: their chemical nature, the possible ways of connection in the space, molecular topology and symmetry, etc.

INFORMATION FOR CHEMICAL COMPOSITION

This is the first information index introduced by Dancoff and Quastler[85] in 1953 as an "information for the kind of atoms in a molecule".

Suppose a molecule has an empirical formula $A_xB_yC_z$. Equs.(6) and (7) allow the total and mean per atom information indices for chemical composition to be calculated in bits:

$$I_{CC} = (x+y+z) \log_2 (x+y+z) - x \log_2 x - y \log_2 y - z \log_2 z \qquad (26)$$

$$\bar{I}_{CC} = - p_x \log_2 p_x - p_y \log_2 p_y - p_z \log_2 p_z \qquad (27),$$

where the probability p_x, p_y, and p_z of a randomly chosen atom to be of the kind A, B, and respectively C, is equal to the relative content of the elements in the molecules: $p_x = x/(x+y+z)$; $p_y = y/(x+y+z)$; $p_z = z/(x+y+z)$. For instance, in the case of C_4H_5Cl these probabilities are $p_C = 0,4$, $p_H = 0,5$, and $p_{Cl} = 0,1$, respectively.

Dancoff and Quastler[85], and Zemanek[86] made use of the information indices for chemical composition in the estimation of the information content of various organisms, including man.

INFORMATION FOR THE WAYS OF LINKING THE ATOMS IN A MOLECULE

In 1955 Morowitz[87] introduced the first information index which takes into account some structural features of molecules. This approach is a combined one. The sum of the information on chemical composition and that of the possible valence bonds between the atoms in a molecule, I_{VB}, is used as a molecular information index:

$$I_{MOR} = I_{CC} + I_{VB} = N \log_2 N - \sum_{i=1}^{k} N_i \log_2 N_i + \sum_{i=1}^{k} N_i \log_2 L_i \qquad (28)$$

In equ.(28) N is the total number of atoms in the molecule, N_i - the number of atoms of the kind i, and L_i - the number of bonds which can be realised by an atom of the kind i. Upon the determination of L_i one deals with the atoms in the molecule as being distributed in cubic cells. Thus, every atom may have no more than six neighbours (fig.5).

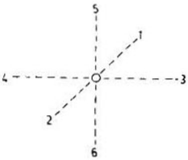


Fig.5. Six possible directions for chemical bonds formation in the Morowitz method

This results in six possible bonds for the elements of valence one ($L_1 = 6$). At higher valences the value of L is greater. In the general case it is specified as the number of combinations of n elements of class r taken r at a time with repetitions:

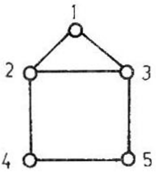
$$L_i \equiv C'_{n,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!} \qquad (29),$$

where $n = 6$, and r is the valence of the atom. The following values of L_i are obtained for elements of a valence 1 to 6: $L_1 = 6$, $L_2 = 21$, $L_3 = 56$, $L_4 = 126$, $L_5 = 256$, and $L_6 = 462$.

The method of Morowitz was widely used in biology for the determination of the information content of organisms in static state, and upon diverse processes as well. Of interest for chemistry are the calculations of information content of a number of classes of organic compounds, as well as the correlation between information indices, heats of formation and entropy of formation of these compounds[88,89].

TOPOLOGICAL INFORMATION INDICES

In the same 1955 Rashevsky[90] introduced another information index to discribe directly the structural features of a molecule on the basis of Graph Theory[91-93]. As is known, the structural formula of chemical compounds can be regarded as graphs (or chemical graphs) whose vertices represent physically indistinguishable atoms and whose edges represent chemical bonds. By the method of Rashevsky the atoms in a molecule are partitioned into classes of topologically equvalent atoms, which are presented by vertices of the same degree in the molecular graph. As a first approximation this makes equivalent all the atoms having the same valence. An additional topological distinction can be introduced taking into account the second, third etc. neighbours. For instance there are three vertices of degree two (1,4,5) and two others of degree three (2,3) in the structure:



Taking into account the second neighbours one discriminates between vertex 1 and the vertices 4 and 5, leaving the latter indistinguishable. Three subsets of topologically equivalent atoms are thus constructed: $\{1\}$, $\{2, 3\}$, and $\{4, 5\}$. A certain amount of structural information corresponds to this distribution of atoms in the molecule. After Rashevsky this kind of information is said to be <u>topological information</u>, I_{top}. Trucco[94] in 1956 specified the definition of the topological equivalence of atoms in the molecule on the basis of the automorphism group of the molecular graph. It is those atoms, which belong to the same orbit of the automorphism group of the graph, that are regarded equivalent, i.e. the atoms which can interchange preserving the adjacency of the graph.

The different chemical nature of the atoms might be additionally taken into account in Rashevsky's method, as in Morowitz's one, by means of the information on chemical composition. Summarizing the latter with the topological information one could obtain the <u>total molecular information content:</u>

$$I_{RASH} = I_{top} + I_{CC} \qquad (30).$$

Making use of the topological information Karreman[95] studied for the first time the information balance of chemical reactions and found it being positive, as well negative, depending essentially on the structure of the reagents and the reaction products.

Rashevsky[90] brought into view the biological importance of the topological information, proceeding from the fact, that the basic vital functions of the organism are, to a great extent, determined by their information content. In a later work[14] he made an estimation of the minimum amount of information, which is necessary for the self-reproduction of an organism. On this basis the fundamental problem for the possibility of spontaneous generation of life on Earth was explored.

In a series of papers Mowshowitz[62,96] developed further the Rashevsky method applying it to the study of the relative complexity of indirected und directed graphs. The topological information is an appropriate measure of complexity, since it depends both on the number of equivalence classes and on their respective cardinalities. The change in this structural index upon various graph operations might be applied to the topological-information modeling of chemical reactions. Exploring this change in detail Mowshowitz found conditions for semi-additivity of information, which are an important contribution to the question of the information balance of chemical reactions. Conditions were also found for a graph to have a prescribed information content, as well as for two graphs to have equal information content.

A second information index for graphs, called <u>chromatic information content</u> , I_{chr}, was also introduced by Mowshowitz[96]. As, in general, there is no unique chromatic decomposition of a graph, it is difficult to specify an information measure which to reflect in a unique way the chromatic structure of a graph. That is why I_{chr} is defined by equs.(6) and (7) as the minimum information over all finite probability schemes constructed from chromatic decompositions having rank equal to the chromatic number of a graph. This information index could be applied to the analysis of some problems in chemistry, first of all to the alternant hydrocarbons whose C-graphs are bicolorable. Exemplifying this information approach one might deal with even alternant hydrocarbons each carbon atom in which carries exactly 1 bit chromatic information (fig.6).

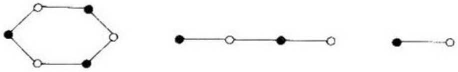


Fig.6. Bicolourable graphs of even alternant hydrocarbons

The number of atoms of each colour in such structures is one and the same. Therefore:

$$N_1 = N_2 = N/2 \quad ; \quad p_1 = p_2 = 1/2 \text{ , and}$$

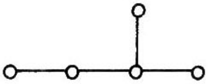
$$\bar{I}_{chr} = - 2 \cdot \frac{1}{2} \log_2 \frac{1}{2} = 1 \text{ bit}$$

Specifying of other information indices for graphs is also possible. As each of them reflects some topological features of structures, the term "topological information" appears to be more general than in the understanding it is advisible for this index to be designated as <u>information on the orbits of the automorphism group of a graph</u>, I_{orb}.

Other topological-information indices can be introduced[97,98] on the basis of the diverse invariants of graphs. This is, for instance, the characteristic polynomial of the graph. One can define an <u>information for polynomial coefficients</u>, p(G,K), of a given graph regarded them as elements of a finite probability scheme. Denoting the sum of polynomial coefficients by z (which for acyclic graphs is identical with the index of Hosoya[99]), the information index thus defined is written out by the equation:

$$I_{pc} = z \log_2 z - \sum_{k=0}^{[N/2]} p(G,k) \log_2 p(G,k) \qquad (31).$$

Here N/2 in Gauss' square brackets is the greatest integer not exceeding the real number in them. For instance, in the case of molecular C-graph of isopentane



$z = 7$, $p(G,0) = 1$, $p(G,1) = 4$, and $p(G,2) = 2$.

Adjacency matrix, incidence matrix and distance matrix (A, I, and D-matrices) are also among the graph invariants. Their matrix elements can be partitioned into subsets of identical elements and a finite probability scheme to be associated with each of these distributions. Thus, a certain information content can be specified for each of these matrices on the basis of equs.(6) and (7).

Adjacency matrix and incidence natrix consist of only zeros and units. Proceeding from this the information on adjacency, as well as the information on incidence of a graph, can be expressed as a function of the total number of graph vertices only:

$$I_{adj} = N^2 \log_2 N^2 - 2(N-1) \log_2 2(N-1) - (N^2-2N+2) \log_2 (N^2-2N+2) \qquad (32)$$

$$I_{inc} = N(N-1) \log_2 N(N-1) - 2(N-1) \log_2 2(N-1) - (N-1)(N-2) \log_2 (N-1)(N-2) \qquad (33)$$

These information indices could be of use in the analysis of some problems connected with the adjadency and incidence of the graphs. Much more sensitive index of the relative complexity of graphs is however the information on distances in the graph:

$$I_D = N^2 \log_2 N^2 - N \log_2 N - \sum_{i=1}^{m} k_i \log_2 k_i \qquad (34),$$

where k_i is an integer that shows how many times the topological distance i occurs in D-matrix. (Due to the symmetry relative to the main diagonal, one might use reduced information indices on the basis of the upper triangular submatrix of A- and D-matrices).

As an illustrative example, the adjacency matrix, incidence matrix and distance matrix of the C-graph of isopentane, as well as the distributions of their matrix elements are shown below:

$$A = \begin{bmatrix} 0&1&0&0&0 \\ 1&0&1&0&0 \\ 0&1&0&1&1 \\ 0&0&1&0&0 \\ 0&0&1&0&0 \end{bmatrix} ; \quad I = \begin{bmatrix} 1&0&0&0 \\ 1&1&0&0 \\ 0&1&1&1 \\ 0&0&1&0 \\ 0&0&1&0 \end{bmatrix} ; \quad D = \begin{bmatrix} 0&1&2&3&3 \\ 1&0&1&2&2 \\ 2&1&0&1&1 \\ 3&2&1&0&2 \\ 3&2&1&2&0 \end{bmatrix}$$

$N_A = 25$, $N_I = 20$, $N_D = 25$; $P_A(8,17)$; $P_I(8,12)$; $P_D(5,8,8,4)$

A series of recent studies of Bonchev, Trinajstić et al[97,98,100,101] has shown the information on distances in the molecular graph as a rather appropriate structural index for the characterization of the branching and cyclicity of molecular skeleton. It excels the topological indices of Hosoya[99], Randic[102], and Wiener[103], as well as the largest eigenvalue in the graph spectrum[104], in isomers discrimination. A combined topological-information approach is devised on this basis that largely facilitates the search of correlations between structure and properties in chemistry. As an example in fig.7 is shown the fairly good correlation[98] between the information on the topological distances in the molecular C-graph and the chromatographic retention indices of mono-, and o-dialkylbenzenes.

More details concerning this approach are given in other reports of this Simposium[105,106]. It has as an essential advantage the possibilities of making predictions, by means of certain topological rules or simple calculations, of the re-

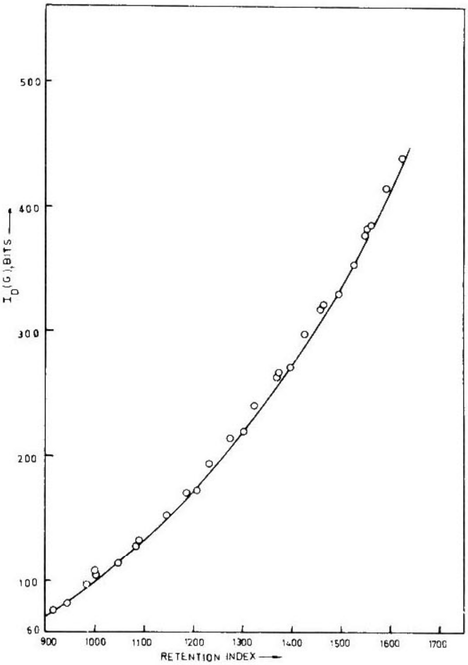


Fig.7. A correlation between the topological information index and the chromatographic retention index of mono-, and o-dialkylbenzenes

lative order of isomers relative to the values of a large number of their properties, including π- and σ-electron energies, heats of formation, boiling points, etc. The topological- information indices could also be of use in the search of efficient ways of coding and conservation of chemical information in computers.

INFORMATION ON MOLECULAR SYMMETRY

The topological approach to the description of molecules has a number of valuable features, and first of all that molecular properties are to a large extent determined by the topology of molecules. On the other hand, the topological description of molecules is not complete, because of the failure to consider the specific molecular geometry-bond lenghts and bond angles. Therefore, it is worthwhile introducing the information on symmetry of molecules[107,108] as an additional structural index. The very important role which the symmetry plays in chemistry[109-111], and especially in quantum mechanical interpretation of atomic and molecular states, spectra, chemical behaviour, etc, gives a sufficient reason for such an index.

The calculations of the information on molecular symmetry is also carried out by equs.(6) and (7) on the basis of the distribution of the atoms in a molecule into classes of atoms equivalent by symmetry. Those atoms are handled as equivalent which interchange upon the operations of the point group of symmetry to which the molecule belongs.

As an illustration transbutadiene molecule is into consideration below (fig.8).

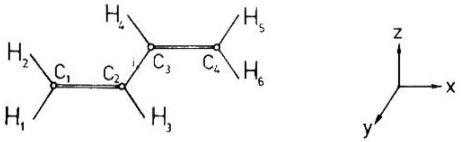


Fig.8. Structure of the butadiene molecule

This molecule belongs to C_{2h} group of symmetry and has as nontrivial elements of symmetry a plane of mirror reflection σ_{xy}, a second order rotation axis C_2^z, and an inversion centre i. The symmetry operations make equivalent the following pairs of atoms:

(C_1,C_4), (C_2,C_3), (H_1,H_6), (H_2,H_5), and (H_3,H_4).

Thus, the distribution on symmetry of all the ten atoms in this molecule is $P_{sym} = (2,2,2,2,2)$.

Analizing the feature of the information index on molecular symmetry one could come to the conclusion that at a constant number of atoms this index increases with the increase in asymmetry. This result is consistent with the Ashby concept of information as a measure of variety in a system, since the great symmetry indicates a great uniformity as well. This structural index excels the topological index of Rashevsky in its ability of discrimination between different molecular conformations.

Information indices for symmetry and topology were compared[108] on the basis of the difference between the automorphism groups of graphs and the point groups of symmetry. It was shown that the two approaches could be modified as to provide the same information content of molecule: $I'_{sym} = I'_{top(orb)}$. This occurs, first, when the topological differences of atoms are taken into account within the classes of atoms with the same chemical nature, and second, when the high local symmetry of the molecular fragments is added to the total molecular symmetry. It should be noted that in this combined method, the idea of equivalent atoms in molecules corresponds in a large number of cases, more to their chemical behaviour than when specified according to the initial methods. These are cases when the equivalence of a certain group of atoms (for example hydrogens in $-CH_3$, $-NH_2$, and other groups) cannot be provided by a total symmetry operation in the molecule, but it is actually ensured by the presence of free rotation of the whole atomic group around a single chemical bond.

The information index on molecular symmetry was applied [112] tocorrelations with various properties of chemical compounds as melting point, vapour pressure, surface tension,etc.

Linear correlations are obtained between this index and absolute entropy for a number of homologous series of organic compounds[108] (fig.9).

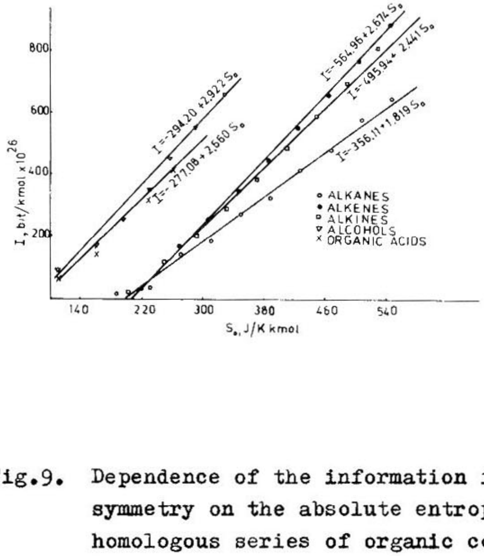


Fig.9. Dependence of the information for molecular symmetry on the absolute entropies for some homologous series of organic compounds

They reveal that the information: entropy ratio is constant in a given homologous series, i.e. every homologous series is characterized by its definite degree of organisation of atoms in molecular structures. The change in the information symmetry which accompanies the change in molecular symmetry was also explored[113]. This differential information index might be of use in the information modeling of chemical processes, such as exchange reactions, molecular rearrangements, cyclization, etc.

The first steps were also made in the development of another information-theoretic approach to molecular symmetry. Subjects of an information analysis in this approach are directly the symmetry groups regarded as mathematical structures, composed of classes of elements of symmetry, irreducible representations and their characters, subgroups, etc. [114,115]. The reflection, which the statistical properties of the point groups thus defined, could have on particular molecular systems is supposed to be studied as a second stage of this approach.

SOME GENERALIZATIONS ON THE MOLECULAR INFORMATION INDICES

The information indices on chemical composition, valence bonds, topology and symmetry of molecules, introduced above, are actually different versions of a general approach to the information study of molecules. All these methods are essentially based on different distributions of atoms in a molecule. The electronic structure of the molecule is not taken into consideration in this approach, or what is equivalent, the atoms in the molecule preserve their individuality. Thus, the information approach "atoms in a molecule" is grounded on similar ideas with the valence bond method in quantum chemistry. In both approaches electrons can come into consideration as electrons of a certain atom. In the information method this could be done specifying the total molecular information content[112] by summation over the atomic and the particular molecular information indices:

$$I_{MOL}^{AM} = I_{MOL}^{i} + \sum_{k=1}^{N} n_k \, I_{AT}^{j} \qquad (35),$$

where i and j represent respectively the different kinds of molecular and atomic information indices, as defined in the preceding chapters, whilst n_k is the number of atoms of the kind k.

This generalized molecular index was also used for correlations with some properties of chemical compounds[112,116]. It is of particular use for small molecules since the variety of atoms in them is too small.

Another information approach is also possible, whose ideas are to some extent close to those of the molecular orbital method in quantum chemistry. In both approaches molecules are regarded as composed of atomic nuclei and electrons, whilst the atoms in molecules are regarded as having completely lost their individuality. This information approach could be termed "nuclei and electrons". Within it the total molecular information content is defined as a sum of the information of nucleus, and respectively electrons, distribution:

$$I^{NE}_{MOL} = I^{NUCL}_{MOL} + I^{EL}_{MOL} \qquad (36).$$

The information indices for chemical composition, topology and symmetry, introduced earlier for the distribution of atoms in a molecule, might be used as a first term in equ.(36). As for the second term, such an information index was quite recently specified by I'Haya et al.[117] for electrons or bond pairs and called "electropy" or, respectively "bondtropy". This index was proposed as a measure of the degree of freedom of choice for electrons or bond pairs in finite space during molecular formation. The calculated values of the "electropy" for a series of linear and cyclic paraffins and monoolefins agree with the observed thermodynamic entropy, heat of formation, heat of compustion, free energy of formation, etc.

To evaluate steric effects, the redundancy of electropy was also considered. In fig.10 the correlation is shown between electropy ε and boiling point (bp) in chain paraffins. Fig.10 shows that ε is a quadratic function of boiling point with a very high correlation coefficient (0,9999):

$$\varepsilon = 0{,}0020(\mathrm{bp})^2 + 0{,}8091\ \mathrm{bp} + 85{,}6720 \qquad (37).$$

In this figure a linear sub-correlation is also shown to exist for isomers (branched paraffins).

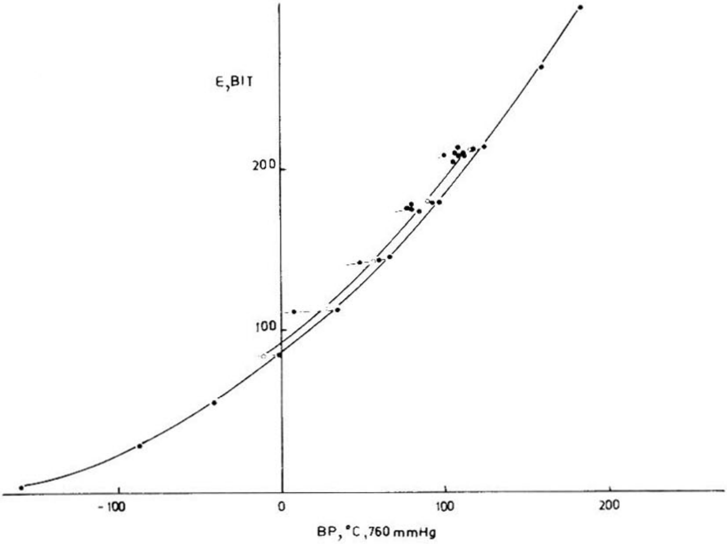


Fig.10. Correlation between electropy $\mathcal{E}$ and boiling point bp in chain paraffins (C_nH_{2n+2}). The mark ⊙ denotes 2-methyl isomer.

One might expect the information approach "nuclei and electrons" to develope fast during the next years, contributing for the broader interplay between Information Theory and quantum mechanics.

Although indirectly associated with the determination of the molecular information content, one may add here the information study of the nodal properties of π -MO (Bonchev, Lickomannov, Trinajstić[118]). An information index is specified in this study by partitioning the interatomic regions in a molecule into nodal and bonding ones, depending on the presence, or respectively absence, of a nodal plane in the space between each pair of neighbouring atoms. The information on nodal properties thus defined was found to be a maximum one at the frontier π-MO (HOMO and LUMO). The equations derived for this index for polyenes and monocyclic aromatic hydrocarbons are a good reason to develop the information approach so that to predict the nodal properties of π-MO in some classes of organic compounds.

Besides the two approaches to the determination of molecular information indices ("atoms in a molecule", and "nuclei and electrons"), atomic groups or molecular fragments can also be involved into consideration as structural elements in the case of more complicated molecules and polymers. Such an approach was developed in connection with the study of the information properties of some biologically important structures, including proteins and DNA-molecule. Branson[119] has calculated the information on aminoacid composition of 26 proteins. The information index of Branson for all the proteins presented in organisms was found to exceed 70% of its maximum value.

Augenstine[120] added to the previous index the information on the configurations of polypeptide chain. Here the exact calculations are hindered by the enormous number of configurations, as well as by the various types of intramolecular bonding decreasing this number. Due to this, estimations of the upper and low limits of this index are only given. The two information quantities, defined above, were applied to the estimation of the quantity of structural information necessary for the normal activity of a protein molecule. In the Augenstine approach a new element appears in the applications of Information Theory to molecules. The information content of a molecule is essentially bonded or potential information. The information on molecular configurations is an active information. It reflects the real discrete states which the molecule could have at normal conditions. Another example of such an approach is the estimation given by Rackow[121] for the information capacity of an asymmetrical carbon atom. Owing to the same probability of existence of D- and L-configurations, this capacity was estimated to 1 bit per molecule (or 1 molbit). A more complete concept of the active information capacity of bioorganic molecules was developed by Zdanov[122]. Within this concept the following discrete states of molecules are regarded as active in informational sence: atomic asymmetry, molecular asymmetry, cis-trans isomerism, stable configurations and tautomerism. The total active information capacity of a molecule is determined by summation over all individual contributions listed above:

$$I_a = \sum_{i=1}^{5} I_{a,i} \tag{38}$$

On this basis the information capacity of different classes of organic compounds is compared.

CONCLUDING REMARKS

It becomes clear from the preceeding presentation that Information Theory is applicable to any chemical structure or process, to which it justaposes a certain information function and a numerical measure. This largely extends the possible applications of Information Theory in chemistry. Some restrictions to the information approach, however, should always be taken into account. Although accurate, the quantitative measures of information are relative and not absolute. They are statistical in character referring to sets but not to individual elements. Information indices are associated with various atomic and molecular properties but the link between them is often rather complicated.

On the other hand, the presence of a large number of information indices for a certain structure could embarrass the explorers. One should however bear in mind that each of the information indices introduced above is a legitimite one. The proper question is not which of these indices are correct and which are not, but whether they are useful and to what extent.

The applicability of the information approach in chemistry is mainly based on the possibilities offered by it for the quantitative analysis of various aspects of chemical structures. The degree of complexity of these structures, as well as their organization and specifity, could be compared in a unique quantitative scale. This provides rigorous studies to be carried out on some general features of the chemical structures as their branching and cyclicity; to examine and compare the degree of organization in different classes of chemical compounds, the specifity of biologically important substances and catalysts, et al.; as well as to approach in a convincing way the question of the degree of similarity or diversity of two chemical objects.

The information approach seems fairly convenient for diverse classificational problems. It is possible in these cases for general information equations to be derived for the main groupings of classified objects (nuclides, groups and periods in the Periodic Table of chemical elements, homologous series of chemical compounds, series of isomeric compounds et al.).

The great power of information methods in discriminating similar structures (isomers, isotopes et al.) is of interest for the computer processing and conservation of chemical information. These methods are fruitful in discriminating between alternative hypotheses or approximations, the latter being of particular interest for quantum chemistry. Alternatively, the possibilities which Information Theory provides for the creation of new hypotheses are limited since this theory describes the mutual dependence of a number of variables, but not the behaviour of each of them.

The relation that exists between structure and properties is another field of successful application of information-theoretic approach in chemistry. The efficiency of this approach was proved on qualitatively different structural levels in chemistry - nuclei, electron shells, molecules, polymers and cristals. Two different directions could be indicated in applying information methods for such aims. Diverse structural rules can be formulated, especially on information-topological basis, which to reflect the mutual influence of two or more structural factors. Chemists can use these simple rules, without any calculations, in predicting the ordering of isomeric compounds relative to a large number of their properties. Quantitative correlations are also possible between the information indices and properties of chemical compounds and elements. Information indices usually gain an advantage over the other structural indices for such correlations since they are capable of expressing the features of chemical structures in full details. High correlations are provided not only with entropy, which is directly associated with information, but also with such important quantities as bonding energy is, although the relation between information and bonding energy is not obvious at all. The properties of individual atoms and molecules are included in these correlations, as well as the properties of large aggregates of them, i.e. such properties which depend on intermolecular and interatomic interactions but not only on their inner structure. In addition, the chemical processes may also be handled by the information approach specifying the change in the information indices upon the interaction, a problem which overlaps to some extent with the problems of graphical reactions.

Most likely, the most attractive feature of information approach is its ability of providing a unified view on various phenomena, as well as a universal language for their description. This approach largely extends the interplay between different scientific disciplines allowing useful analogies, as well as common laws to be devised. The contemporary science tends to unity and Information Theory is one of the promising ways to it. The chemical applications of Information Theory have their important place within this tendency, and one should expect their importance to raise in the future.

LITERATURE

1. N.Wiener, Cybernetics (Wiley, New York, 1948).
2. W.Ashby, An Introduction to Cybernetics (Wiley, New York, 1956).
3. V.Glushkov, On Cybernetics as a Science, in Cybernetics, Thinking,Life (russ.), (Misl, Moscow, 1964, p.53).
4. C.Shannon and W. Weaver, Mathematical Theory of Communication (University of Illinois, Urbana, 1949).
5. L.Boltzmann, Vorlesungen über Gastheorie (Leipzig, 1896).
6. Szilard, Z.Phys., No.53 (1929).
7. L.Brillouin, Science and Information Theory (Academic, New York, 1956).
8. L.Brillouin, Scientific Uncertainty and Information (Academic, New York, 1964).
9. D.Bonchev and G.Lickomannov, Information Theory, Thermodynamics, and Statistics (A Selected Bibliography), Math. Chem. 3,269(1977).
10. N.I.Kobozev, A Study on the Thermodynamics of Information and Thinking Processes (russ.), (Moscow University, 1971).
11. Essays on the Use of Information Theory in Biology, ed.by H.Quastler (University of Illinois, Urbana, 1953).
12. P.G.Seybold, Intern.J.Quantum Chem.: Quantum Biology Symp. No.3,39(1976).
13. M.Hasegawa and T.A.Yano, Origins Life 6,219(1975).
14. N.Rashevsky, Bull.Math.Biophysics 22,351(1960); C.Yockey and P.Hubert, J.Theor.Biol. 67,377(1977).
15. K.S.Trincher, Biology and Information (russ.), (Nauka, Moscow, 1964).
16. A.I.Bykhovskii, Math.Biosci. 3,353(1968).
17. A.D.Ursul, Filossofskie Nauki No.2,43(1966), (russ.).
18. D.Bonchev, Biology and Chemistry 13,4(1970), (bulg.).
19. E.Schroedinger, What is Life? (University Press, Cambrigde, England, 1944).
20. I.I.Brekhman, Man and Biologically Active Substances (russ.), (Nauka, Moscow, 1976).

21. R.B.Bernstein and R.D.Levine, J.Chem.Phys. 57,434(1972).
22. A.Ben-Shaul, R.D.Levine, and R.B.Bernstein, J.Chem.Phys. 57,5427(1972).
23. R.D.Levine and R.B.Berstein, Acc.Chem.Research 7,393(1974).
24. R.B.Bernstein and R.D.Levine, Adv.At.Mol.Phys. 11,215 (1975).
25. N.Agmon and R.D.Levine, Chem.Phys.Lett. 52,197(1977).
26. R.K.Huddleston and E.Weitz, J.Chem.Phys. 66,1740(1977).
27. E.Ludena and V.Amzel, J.Chem.Phys. 52,5293(1970).
28. C.Aslangul, R.Constanciel, R.Daudel and P.Kottis, Adv. Quant.Chem. 6,93(1972).
29. R.Daudel, R.F.Bader, M.E.Stefens and D.S.Borett, Can.J. Chem. 52,1310(1974); ibid. 52,3077(1974).
30. C.Aslangul, R.Constanciel, R.Daudel, L.Esnault and E.Ludena, Int.J.Quant.Chem. 8,499(1974).
31. N.I.Kobozev, B.V.Strakhov and A.M.Rubashov, Zh.Fiz.Khim. 45,86(1971); ibid. 45,375(1971).
32. N.I.Kobozev and B.V.Strakhov, Osn.Predvideniya Katal. Deistviya, in: Ya.T.Eidus (Ed.), Tr.Mezdunar.Kongr.Katal., 4th 1963 (Nauka, Moscow, 1970) p.307.
33. N.I.Kobozev and B.V.Strakhov, Proc.Int.Congr.Catal., 4th 1968 (Akad.Kiado, Budapest, 1971) p.332.
34. V.S.Beskov and N.N.Solyakina, Tr.N.I.I.Proekt.In-ta Azot. Prom-sti i Produktov Organ.Sinteza 36,49(1975).
35. B.Rackow, Z.Chem 7,399(1967); ibid. 7,472(1967); ibid. 8,33(1968); ibid. 9,318(1969).
36. T.G.Petrov, Protsessi Rosta Krist.Plenok Poluprov., in L.A.Aleksandrov (Ed.), Tr.Simp., 2nd 1969 (Sib.Otd.Akad. Nauk SSSR, Novosibirsk, 1970) p.61.
37. D.H.Andrews and M.L.Boss, Yale Sci., 45,2,6,9,11(1971).
38. B.V.Andrew, Nature 202,1206(1964).
39. G.A.Bulkin and K.B.Gatalin, Geol.Zh.(russ.) 36,76(1976).
40. T.G.Petrov, Dokl.Akad.Nauk SSSR 191,924(1970).
41. T.G.Petrov, Vest.Leningrad.Univ., Geol., Geogr. 3,30 (1971).

42. E.Eckschlager, Collect.Czech.Chem.Communs. 36,3016(1971); ibid. 37,137(1972); ibid. 37,1486(1972); ibid. 38,1330 (1973).
43. G.Gottschalk, Frezenius's Anal.Chem. 258,1(1972).
44. Yu.I.Belyaev and T.A.Koveshnikova, Zh.Anal.Khim. 27,429 (1972).
45. K.Doerffel, Chem.Techn. (Leipzig) 25,94(1973).
46. St.Grys, Frezenius's Anal.Chem. 273,177(1975).
47. H.Malissa, J.Rendl and I.L.Marr, Talanta 22,597(1975).
48. B.Rackov, Z.Chem. 4,109(1964); ibid. 4,155(1964); ibid. 5,67(1965).
49. F.Dupuis and A.Dijkstra, Anal.Chem. 47,379(1975).
50. Yu.I.Stakheev, Zavodsk.Lab. 28,831(1962).
51. H.G.Bartel, Z.Chem. 15,62(1975).
52. A.I.Khinchin, Mathematical Foundations of Information Theory (Dover, New York, 1957).
53. E.T.Jaynes, Phys.Rev. 106,620(1957); ibid. 108,171(1957).
54. M.Tribus, Thermostatics and Thermodynamics (Van Nostrand, New York, 1964).
55. A.Katz, Principles of Statistical Mechanics - The Information Theory Approach (Freeman, San Francisco, 1967).
56. R.S.Ingarden and K.Urbanik, Bull.Acad.Sci.Polon., Ser. sci math. 9,313(1961).
57. R.S.Ingarden and K.Urbanik, Colloq.Math. 9,131(1962).
58. A.N.Kolmogorov, Problemi peredachi informatsii 1,3(1965).
59. A.N.Kolmogorov, Vestnik AN SSSR, No5(1965).
60. A.N.Kolmogorov, Problemi peredachi informatsii 5,3(1969).
61. A.Renyi, Rev.Intern.Statist. 3,1(1965).
62. A.Mowshowitz, Bull.Math.Biophys. 30,175(1968).
63. D.Bonchev and T.Peev, Jahrb.Hochsch.Chem.Techn.Burgas 10,561(1973).
64. M.Göppert-Mayer, Nuclear Shell Structure (Mc Graw Hill, New York, 1961).
65. T.Peev, D.Bonchev, A.Dimitrov and B.Rousseva, Ann.Ecol. Super.Phys.Techn. 11,87(1974).
66. T.Peev, D.Bonchev, B.Rousseva and A. Dimitrov, Jahrb.

Hochsch.Chem.Techn.Burgas 9,73(1972).

67. D.Bonchev, T.Peev and B.Rousseva, Math.Chem. 2,123(1976).
68. C.Weizsäcker, Z.Phys. 96,431(1935).
69. Chart of the Nuclides, Sec.Edition (Gersbach, München, 1963).
70. B.Rousseva and D.Bonchev, Math.Chem. (submitted).
71. V.Kravtzov, Atomic Masses and Nuclear Binding Energies (russ.), (Atomisdat, Moscow, 1974).
72. E.Clementi and D.L.Raimondi, J.Chem.Phys. 38,2686(1963); E.Clementi, D.L.Raimondi and W.P.Reinhardt, J.Chem.Phys. 47,1300(1967).
73. D.Bonchev, V.Kamenska and D.Kamenski, Mh.Chem. 108,477 (1977).
74. D.Bonchev and V.Kamenska, Croat.Chim.Acta (in press).
75. D.Dimov and D.Bonchev, Math.Chem. 2,111(1976).
76. D.Bonchev, V.Kamenska and C.Tashkova, Math.Chem. 2,117 (1976).
77. D.Bonchev and V.Kamenska, Mh.Chem. 109,551(1978).
78. V.Klechkovski, Distribution of Atomic Electrons and the Rule of Subsequent Filling of the (n+l)-Groups (russ.), (Atomisdat, Moscow, 1968).
79. G.Villar, J.Inorg.Nucl.Chem. 28,25(1966).
80. D.Bonchev and V.Kamenska, Mh.Chem. (submitted).
81. J.Andrade e Silva and G.Lochak, C.r.Acad.Sci. Nr.25,4260 (1962).
82. L.de Broglie, C.r.Acad.sci. Nr.6,1052(1962).
83. D.Bonchev and V.Kamenska (This Simposium).
84. D.Bonchev, J.Math.Phys. (submitted).
85. S.M.Dancoff and H.Quastler, In Essays on the Use of Information Theory in Biology, ed.by H.Quastler (University of Illinois, Urbana, 1953).
86. H.Zemanek, Elementare Informationstheorie (Oldenbourg, Wien, 1959).
87. H.Morowitz, Bull.Math.Biophys. 17,81(1955).
88. M.Valentinuzzi and M.E.Valentinuzzi, Anal.Soc.Cientif. Argentina 1(1962).

89. M.Valentinuzzi and M.E.Valentinuzzi, Bull.Math.Biophys. 25,11(1963).
90. N.Rashevsky, Bull.Math.Biophys. 17,229(1955).
91. F.Harary, Graph Theory (Addison-Wesley, Reading, MA, 1971), 2nd ed.
92. M.Behzad and G.Chartrand, Introduction to the Theory of Graphs (Allyn and Bacon, Boston, 1972).
93. Chemical Application of Graph Theory, ed.by A.T.Balaban (Academic, London, 1976).
94. E.Trucco, Bull.Math.Biophys. 18,129(1956); ibid. 18,237 (1956).
95. G.Karreman, Bull.Math.Biophys. 17,279(1955).
96. A.Mowshowitz, Bull.Math.Biophys. 30,225(1968); ibid. 30,387(1968); ibid. 30,533(1968).
97. D.Bonchev and N.Trinajstić, J.Chem.Phys. 67,4517(1977).
98. D.Bonchev and N.Trinajstić, Intern.J.QuantumChem. (in press).
99. H.Hosoya, Bull.Chem.Soc.Japan 44,2332(1971; J.Chem.Soc. 12,181(1972).
100. D.Bonchev, O.Mekenian and N.Trinajstić (to be submitted).
101. O.Mekenian, D.Bonchev and N.Trinajstić (to be submitted).
102. M.Randić, J.Am.Chem.Soc. 97,6609(1975).
103. H.Wiener, J.Am.Chem.Soc. 69,17(1947).
104. L.Lovász and J.Pelikan, PeriodMath.Hung. 3,175(1973).
105. O.Mekenian, D.Bonchev and N.Trinajstić (This Simposium).
106. D.Bonchev, J.Knopp and N.Trinajstić (This Simposium).
107. D.Bonchev, D.Kamenski and V.Kamenska, Jahrb.Hochsch.Chem. Techn.Burgas 10,437(1973).
108. D.Bonchev, D.Kamenski and. V.Kamenska, Bull.Math.Biol. 38,119(1976).
109. M.Hammermesh, Group Theory and its Application to Physical Problems (Addison-Wesley, London, 1964).
110. R.Hochstrasser, Molecular Aspects of Symmetry (Benjamin, New York, 1966).
111. F.A.Cotton, Chemical Applications of Group Theory, 2nd ed. (Wiley-Interscience, New York, 1971).
112. D.Bonchev and O.Mekenian, Jahrb.Hochsch.Chem.Techn.Burgas

10,423(1973).
113. D.Dimov and D.Bonchev, Bull.Math.Biol. (submitted).
114. D.Bonchev and G.Lickomannov, Math.Chem. 3,3(1977).
115. D.Bonchev and G.Lickomannov, Intern.J.Quantum Chem. (submitted).
116. D.Bonchev, C.Tashkova and R.Ljutskanova, C.r.Bulg.Acad. Sci. 28,225(1975).
117. W.T.Yee, K.Sakamoto and Y.J.I'Haya, Rep. Univ.Electro-Comm. 27,53(1976); ibid. 27,227(1977).
118. D.Bonchev, G.Lickomannov and N.Trinajstić, Acta Chim. Hung. (submitted).
119. H.Branson, in Information Theory in Biology, ed.by H.Quastler (University of Illinois, Urbana, 1953),84.
120. L.Augenstine, in Information Theory in Biology, ed.by H.Quastler (University of Illinois, Urbana, 1953), 105, 119.
121. B.Rackow, Z.Chem. 3,268(1963).
122. Yn.A.Zdanov, Biofiz. (russ.) 12,715(1967).