

Encoding the Core Electrons with Graph Concepts

Lionello Pogliani*

Dipartimento di Chimica, Università della Calabria, 87030 Rende (CS), Italy

Received June 21, 2003

The core electron problem of atoms in chemical graph studies has always been considered as a minor problem. Usually, chemical graphs had to encode just a small set of second row atoms, i.e., C, N, O, and F, thus, graph and, in some cases, pseudograph concepts were enough to “graph” encode the molecules at hand. Molecular connectivity theory, together with its side-branch the electrotopological state, introduced two “ad hoc” algorithms for the core electrons of higher-row atoms based, mainly, on quantum concepts alike. Recently, complete graphs, and, especially, odd complete graphs have been introduced to encode the core electrons of higher-row atoms. By the aid of these types of graphs a double-valued algorithm has been proposed for the valence delta, δ^v , of any type of atoms of the periodic table with a principal quantum number $n \geq 2$. The new algorithm is centered on an invariant suggested by the *hand-shaking* theorem, and the values it gives rise to parallel in some way the values derived by the aid of the two old “quantum” algorithms. A thorough comparative analysis of the newly proposed algorithms has been undertaken for atoms of the group 1A–7A of the periodic table. This comparative study includes the electronegativity, the size of the atoms, the first ionization energy, and the electron affinity. The given algorithm has also been tested with sequential complete graphs, while the even complete graphs give rise to conceptual difficulties. QSAR/QSPR studies do not show a clear-cut preference for any of the two values the algorithm gives rise to, even if recent results seem to prefer one of the two values.

INTRODUCTION

Molecular connectivity theory with its side-branch the Electrotopological State, E_s , is mainly based on graph and pseudograph concepts. The encoding of a hydrogen-suppressed molecule is achieved by the aid of these two types of concepts, which are self-sufficient if the hydrogen-suppressed molecule is made up of second row atoms only. In fact, most of the studied molecules are made up of C, N, O, and F.^{1–3} Molecular connectivity, to solve the problem of the core electrons with atoms with principal quantum number $n > 2$ (higher-row atoms), developed two different algorithms for its basic parameter, the valence delta, δ^v . Normally, the higher-row atoms are the atoms belonging to the groups 1A–7A of the periodic table of the elements. The first proposed algorithm for δ^v , which for second-row atoms can be derived by the aid of pseudograph concepts only, is nearly as old as molecular connectivity and is based on quantum concepts

$$\delta^v = (Z^v - h)/(Z - Z^v - 1) \quad (1)$$

Here Z is the atomic number, Z^v is the number of valence electrons, and h is the number of hydrogens. For second row atoms $\delta^v = Z^v - h \equiv \delta^v(\text{ps})$, where $\delta^v(\text{ps})$ is the valence delta value that can be derived by the aid of pseudograph concepts only, as $Z = Z^v + 2$. Actually the all-purposes $\delta^v(Z^v)$ of eq 1, which is valid for atoms of any row could be rewritten as

$$\delta^v(Z^v) = \delta^v(\text{ps})/(Z - Z^v - 1) \quad (2)$$

The second type of algorithm for δ^v has recently been proposed within the frame of the E_s state concept.² This branch of molecular connectivity develops two new indices apt to encode specific atomic characteristics of atoms in molecules, of which one is the intrinsic state, I_s , index

$$I = [(2/n)^2 \delta^v + 1]/\delta \quad (3)$$

Here, n is the principal quantum number, and $\delta^v = Z^v - h$. As δ^v can be derived from the pseudograph of a molecule we prefer to name it as $\delta^v(\text{ps})$, while δ is a parameter that can be derived from the corresponding graph of a molecule. Thus, the following all-purposes $\delta^v(E_s)$ can here be defined for atoms of any row (clearly, now, $I = [\delta^v + 1]/\delta$)

$$\delta^v(E_s) = (2/n)^2 \delta^v(\text{ps}) \quad (4)$$

Here the principal quantum number n encodes the contribution of the core electrons. For $n = 2$ we have a result already obtained with eq 1, i.e., $\delta^v = \delta^v(\text{ps})$, while for $n > 2$ the two algorithms diverge, as can be easily seen from Figure 1. In Figure 1 the δ^v values defined within the frame of Z/Z^v parameters, i.e., $\delta^v = f(Z^v)$ (■ (black)), and defined within the frame of the Electrotopological State parameters, i.e., $\delta^v = f(E_s)$ (▲ (red)), have been plotted as a function of the atomic number for the atoms belonging to the groups 1A to 7A (see Table 1). As some atoms can have more than one $\delta^v(\text{ps})$ value, i.e., C whose $\delta^v(\text{ps})$ can be 1, 2, 3, and 4, only the highest values for each atom have been chosen here. In this figure is clearly noticeable (i) the agreement between

* Corresponding author e-mail: lionp@unical.it. On sabbatical leave, present address: Facultad de Farmacia, Dept. de Química Física, Universidad de Valencia, 46100 Burjassot, Valencia, Spain.

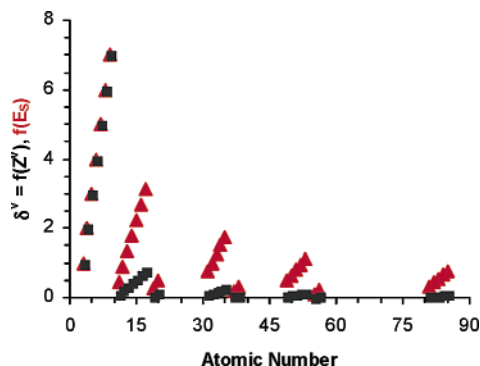


Figure 1. The δ^v once as a function of the Z^v parameter (old definition, eq 1, \blacksquare), $\delta^v = f(Z^v)$ and the other time defined within the frame of the Electrotopological state index (eq 2, \blacktriangle), $\delta^v = f(E_s)$, vs the atomic number for the atoms of groups 1A–7A.

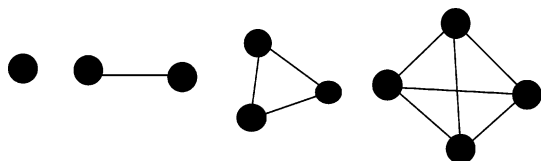


Figure 2. The K_1 , K_2 , K_3 , and K_4 (from left to right) complete graphs.

the two δ^v for $n = 2$ (n = principal quantum number), (ii) the divergence of their values for $n > 2$, and (iii) the stronger variation, with the atomic number, of the $\delta^v = f(E_s)$ (\blacktriangle (red)).

METHOD

The Graph Conjecture. Recently a graph conjecture for the core electrons of the atoms with $n \geq 2$ has been introduced to deepen the graph character of molecular connectivity theory and get rid of a δ^v partially based on quantum concepts.⁴ This conjecture, which is based on the concept of a complete graph, has been successfully tested with several classes of compounds.^{5,6} A graph is complete if every two of its vertices are adjacent. A complete graph of order p is denoted by K_p and is r -regular, where r is its regularity and $r = p - 1$. A graph is r -regular if it has all its vertices with the same degree, i.e., all complete graphs are regular but not all regular graphs are complete. In Figure 2 are displayed the K_1 , K_2 , K_3 , and K_4 complete graphs. There is also a type of complete graph, which cannot be displayed, i.e., the null graph, K_0 , which possesses no points and no edges (the nothing, with negative regularity!!).⁷ The algorithm that has successfully been tested, and which is based

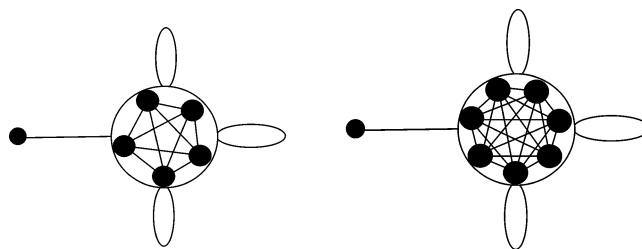


Figure 3. The pseudograph-odd-complete-graphs of the hydrogen-suppressed CH_3Br and CH_3I . Odd complete graphs are shown in a kind of zoom of the vertices of Br and I. A carbon atom is here encoded by a K_1 vertex.

on two important parameters of complete graphs, p and r is the following

$$\delta^v = q \cdot \delta^v(\text{ps}) / [p \cdot r + 1] \quad (5)$$

where p is the order of the complete graph and r its regularity. Parameter $p \cdot r$ in eq 5 has a particular importance in graph theory. This parameter equals the sum of all vertex degrees in complete graphs, and it is related to a famous theorem in graph theory, the “handshaking theorem” that says that this sum, for every type of graph and pseudograph, equals twice the number of connections.⁸ Actually, only one type of complete graph was successful for practical and theoretical reasons, the odd complete graphs with $p = 1, 3, 5, 7, \dots$ and $r = 0, 2, 4, 6, \dots$. Sequential complete graphs with $p = 1, 2, 3, 4, \dots$ gave rise to poor results up to date. Even complete graphs, instead, with $p = 2, 4, 6, \dots$ give rise to a series of conceptual difficulties, as they would oblige to redefine all those graphs made up of second row atoms, i.e., made up, practically, of K_1 vertices. In fact, here either a K_0 graph or a K_2 graph should be introduced for the second row atoms. Now, K_0 possesses no points and no edges, and for this reason should be discarded for chemical graphs. K_2 has two vertices and an edge, and this choice would oblige to redefine all the chemical graphs made up of second row atoms, a choice that we prefer to avoid. The given odd but also sequential complete graph algorithm allows preserving the single-vertex K_1 representation for all those second row atoms of the hydrogen-suppressed chemical graphs and pseudographs.

Equation 5 has a parameter that needs to be defined, the q parameter. Studied models show that in some cases $q = p$, and then the δ^v values are rather similar to the $\delta^v = (2/n)^2 \delta^v(\text{ps})$ values of the E_s algorithm, and in some other cases, $q = 1$. In this last case the δ^v values are rather similar to the values obtained with $\delta^v = \delta^v(\text{ps}) / (Z - Z^v - 1)$. In Figure 3

Table 1. Electronegativities and Atomic Radii (pm) of Thirty Elements* Belonging to the Main Groups 1A–7A and to the Periods $n = 2$ –6^a

	1A	2A	3A	4A	5A	6A	7A
2 (K_1)	Li (3) 1.0; 152	Be (4) 1.5; 112	B (5) 2.0; 98	C (6) 2.5; 91	N (7) 3.0; 92	O (8) 3.5; 73	F (9) 4.0; 72
3 (K_3)	Na (11) 0.9; 186	Mg (12) 1.2; 160	Al (13) 1.5; 143	Si (14) 1.8; 132	P (15) 2.1; 128	S (16) 2.5; 127	Cl (17) 3.0; 99
4 (K_5)	K (19) 0.8; 227	Ca (20) 1.0; 197	Ga (31) 1.6; 135	Ge (32) 1.8; 137	As (33) 2.0; 139	Se (34) 2.4; 140	Br (35) 2.8; 114
5 (K_7)	Rb (37) 0.8; 248	Sr (38) 1.0; 215	In (49) 1.7; 166	Sn (50) 1.8; 162	Sb (51) 1.9; 159	Te (52) 2.1; 160	I (53) 2.5; 133
6 (K_9)	Cs (55) 0.7; 265	Ba (56) 0.9; 222	Tl (81) 1.8; 171	Pb (82) 1.9; 175	Bi (83) 1.9; 170	Po (84) 2.0; 164	At (85) 2.2; 142

^a For each atom in parentheses is the atomic number, Z . Along the first column, in parentheses, are the K_p odd complete graphs.

are shown the hydrogen-suppressed pseudographs-odd-complete-graphs for $\text{CH}_3\text{-Br}$ and $\text{CH}_3\text{-I}$. While the odd complete graph for the carbon atom is just a K_1 vertex, the complete graphs for Br and I are K_5 and K_7 complete graphs, respectively (Figure 3 shows a blow up of the Br and I vertices). Let us denote algorithm (5) for $p = q$ by K_{pp} and for $q = 1$ by K_p . In Figure 4 are shown the nearly symmetric relations, which hold between $\delta^v(K_p)$ and $\delta^v(Z^v)$ (top) and between $\delta^v(K_{pp})$ and $\delta^v(E_s)$ (bottom). This nearly symmetric relation does not hold for p sequential, i.e., for $p = 1, 2, 3, 4, \dots$. Actually, the two sets of values for p sequential, the K_{pp} and the K_p values, are more similar to each other than the two sets of values for p odd.

The Basis Indices. The following subsets of basis indices will here be used for model purposes, the molecular connectivity indices, the molecular pseudoconnectivity indices and the dual indices,^{1-3, 9-11}

$$\begin{aligned}\{\chi\} &= \{\{D, \chi, {}^1\chi, \chi_t, D^v, \chi^v, {}^1\chi^v, \chi_t^v\}\} \\ \{\psi\} &= \{\psi_I, {}^0\psi_I, {}^1\psi_I, {}^T\psi_I, {}^S\psi_E, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\} \\ \{\beta_d\} &= \\ &\{\chi_d, {}^1\chi_d, {}^1\chi_s, {}^0\chi_d, {}^1\chi^v, {}^1\chi_s^v, {}^0\psi_{Id}, {}^1\psi_{Id}, {}^1\psi_{Is}, {}^0\psi_{Ed}, {}^1\psi_{Ed}, {}^1\psi_{Es}\} \quad (6)\end{aligned}$$

i.e., $\{\beta\} = \{\{\chi\}, \{\psi\}, \{\beta_d\}\}$. Basis χ and ψ indices are formally similar, in fact

$$D = \sum_i \delta_i {}^S\psi_I = \sum_i I_i \quad (7)$$

$${}^0\chi = \sum_i (\delta_i)^{-0.5} {}^0\psi_I = \sum_i (I_i)^{-0.5} \quad (8)$$

$${}^1\chi = \sum (\delta_i \delta_j)^{-0.5} {}^1\psi_I = \sum (I_i I_j)^{-0.5} \quad (9)$$

$$\chi_t = (\Pi \delta_i)^{-0.5} {}^T\psi_I = (\Pi I_i)^{-0.5} \quad (10)$$

Index χ_t (and χ^v) is the total molecular connectivity index, and it has its ψ counterpart in the total molecular pseudoconnectivity index, ${}^T\psi_I$ (and ${}^T\psi_E$). Sums in eqs 7 and 8 as well as products (Π) in eq 10 are taken over all vertices of the hydrogen-suppressed chemical graph. Sums in eq 9 are over all edges of the chemical graph (σ bonds in a molecule). Replacing δ with δ^v the subset of valence χ^v indices, $\{D^v, {}^0\chi^v, {}^1\chi^v, \chi_t^v\}$ for a hydrogen-suppressed chemical pseudograph is obtained. Replacing I_i with S_i the ψ_E subset $\{{}^S\psi_E, {}^0\psi_E, {}^1\psi_E, {}^T\psi_E\}$ is obtained. Superscripts S and T stand for sum and total, while the other sub- and superscripts follow the established denomination for χ indices.

The dual basis indices are defined in the following way¹¹

$${}^0\chi_d = (-0.5)^N \Pi_i (\delta_i) {}^0\psi_{Id} = (-0.5)^N \Pi_i (I_i) \quad (11)$$

$$\begin{aligned}{}^1\chi_d &= (-0.5)^{(N+\mu-1)} \Pi (\delta_i + \delta_j) \psi_{Id} = \\ &(-0.5)^{(N+\mu-1)} \Pi (I_i + I_j) \quad (12)\end{aligned}$$

$${}^1\chi_s = \Pi (\delta_i + \delta_j)^{-0.5} {}^1\psi_{Is} = \Pi (I_i + I_j)^{-0.5} \quad (13)$$

Replacing, in these expressions δ by δ^v and I_i by S_i the corresponding χ^v valence dual and ψ_E dual indices are obtained. Exponent μ in eq 12 is the cyclomatic number. It

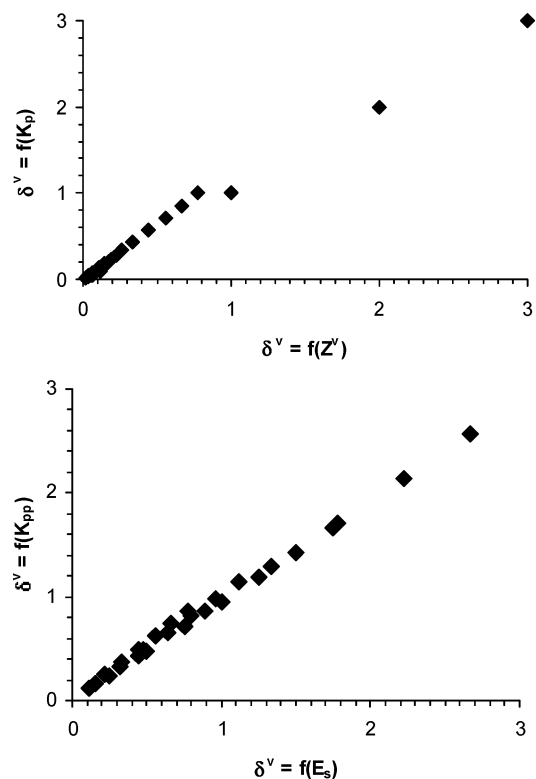


Figure 4. **Top** — the $\delta^v(K_p)$ values with p odd vs the $\delta^v = f(Z^v)$ values and **bottom** — the $\delta^v(K_{pp})$ values with p odd vs the $\delta^v = f(E_s)$ values (for p sequential see text).

indicates the number of cycles of a chemical graph, and it is equal to the minimal number of edges necessary to be removed in order to convert the (poly)cyclic graph to an acyclic subgraph.

Basis ψ indices are indirectly related to the δ^v number, defined in eq 5, through the I-State (ψ_I subset) and S-State (ψ_E subset) indices: ${}^2I = (\delta^v + 1)/\delta$, $S = I + \Sigma \Delta I$, and $\Delta I = (I_i - I_j)/r_{ij}^2$. Here, r_{ij} counts the atoms in the minimum path length separating two atoms i and j , which is equal to the usual graph distance, $d_{ij} + 1$. Factor $\Sigma \Delta I$ incorporates the information about the influence of the remainder of the molecular environment, and, as it can be negative, S can also be negative. To avoid imaginary ψ_E values, every S value of our classes of compounds (as some atoms have $S < 0$), has been rescaled to the S value in SiF_4 for the studied organic compounds, i.e., $S(\text{Si}) = -6.611$. The S values of the studied inorganic compounds has been rescaled to the S value of Ba in BaF_2 , i.e., $S[\text{Ba}(\text{BaF}_2)] = -3.083$. The rescaling procedure has minor effects on the model.¹⁰

With these basis indices it is possible to construct a series of higher-order descriptors, known as terms, $X = f(\chi)$, $Y = f(\psi)$, $Z = f(X, Y, \beta)$, $Z' = f(Z, \beta_d)$.^{3,12} The quality of any description is under the control of several statistics, r , s , the Fisher ratio, F , and $Q = r/s$, which measures for a property the growing quality of successive descriptors. Further, to have a direct idea of how much s improves along different descriptors, $s_R = s_0/s_i$ has been used, where s_0 is the s value for the best single-basis-index, and s_i is the s value of the other descriptors. To measure the utility of the regression parameters, $u_i = c_i/e_i$ has been used, where c_i is the value of a parameter of the linear regression, and e_i is its error. These last values as well as the parameters of the best linear

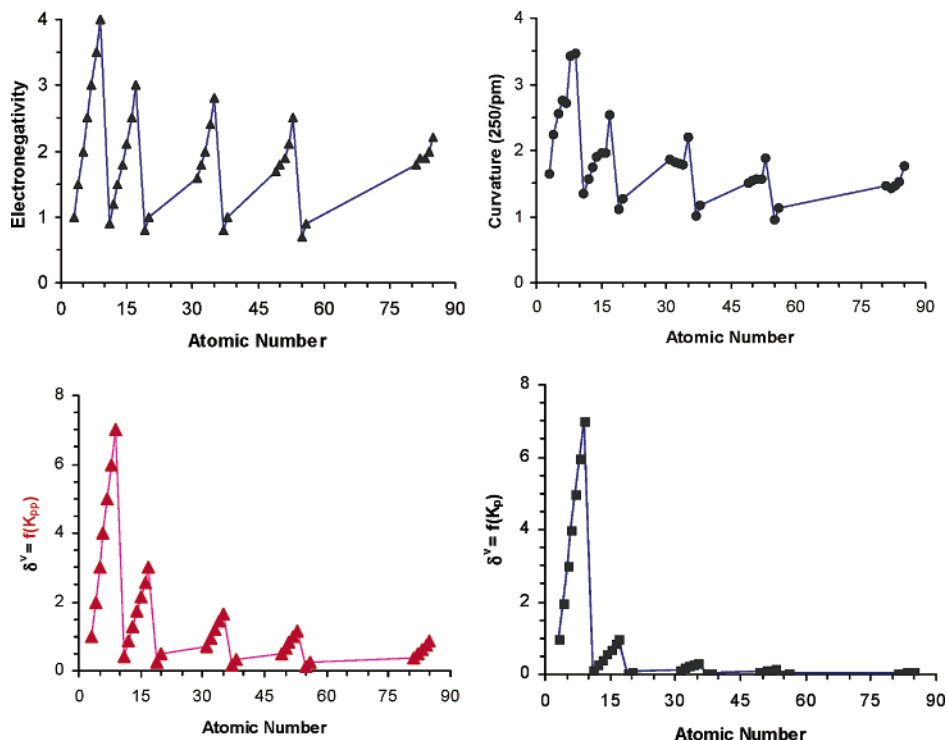


Figure 5. Top – (left) the electronegativity of 1A–7A atoms vs their atomic number; (right) the curvature as 250/pm of 1A–7A atoms vs their atomic number and bottom – (left) the $\delta^v(K_{pp})$ values with p odd vs the atomic number for atoms 1A–7A, (right) the $\delta^v(K_p)$ values with p odd vs the atomic number for atoms 1A–7A.

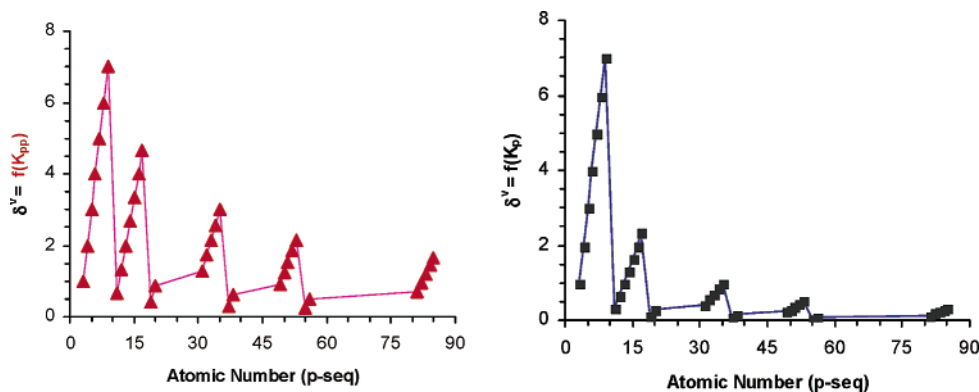


Figure 6. Left – the $\delta^v(K_{pp})$ values with p sequential vs the atomic number of atoms 1A–7A and right – the $\delta^v(K_p)$ values with p sequential vs the atomic number of atoms 1A–7A.

regression are collected into the form of a vector \mathbf{u} and \mathbf{C} , respectively (for more details see refs 3–5).

RESULTS AND DISCUSSION

The Electronegativity, the Curvature, the First Ionization Energy, the Electron Affinity, and δ^v . In Table 1 are collected the electronegativity values (first value in the second row of each element) and the atomic radii in pm (second values in the second row of each element). These values have been taken from Chang.¹³ Let us now see how the $\delta^v(K_p)$ and $\delta^v(K_{pp})$ with p odd relate to the electronegativity and to the curvature of the atoms, where the curvature equals the reciprocal of the radii. In Figure 5 are reported the electronegativity (left) and the curvature (right) as a function of the atomic number. Curvature has, here, been normalized to show a similarity with the electronegativity values. At the bottom of the figure are shown the $\delta^v(K_{pp})$ (left) and $\delta^v(K_p)$ vs the atomic number. The best agreement

among these data is clearly the one among the $\delta^v(K_{pp})$ data, with the electronegativity and the curvature. This is especially true for the atom belonging to $n = 3, 4, 5$, and 6. In fact, the agreement with the absolute electronegativity values and the normalized curvature values is astounding. Figure 6 is obtained with p sequential; here a more detailed analysis shows a discrepancy at the level of $n = 3$ for K_{pp} but a better agreement for K_p . Further the $K_{pp}(\text{p-seq})$ and $K_p(\text{p-seq})$ values are closer to each other than those corresponding to the $K_{pp}(\text{p-odd})$ and $K_p(\text{p-odd})$ values, i.e., the odd complete case guarantees a better resolution between the K_{pp} and the K_p values.

In Table 2 are collected the first ionization energy (first value in the second row of each element) and the electronic affinity (second value in the second row of each element) taken from Atkins.¹⁴ Ionization energies and the electronic affinity have been plotted vs the atomic number and are shown in Figure 7. The $K_{pp}(\text{p-seq})$ values are the set of

Table 2. First Ionization Energy E_1 (kJ mol⁻¹) and Electron Affinities E_a (kJ mol⁻¹) of the Main Groups 1A–7A of the Periods $n = 2–6^a$

	1A	2A	3A	4A	5A	6A	7A
2 (K₁)	Li (3) 513; 59.8	Be (4) 899; -18	B (5) 801; 23	C (6) 1086; 123	N (7) 1402; -7	O (8) 1314; 141	F (9) 1681; 322
3 (K₃)	Na (11) 496; 52.9	Mg (12) 738; -21	Al (13) 577; 44	Si (14) 787; 134	P (15) 1012; 71.7	S (16) 1000; 200	Cl (17) 1251; 349
4 (K₅)	K (19) 419; 48.3	Ca (20) 590; -186	Ga (31) 579; 36	Ge (32) 762; 116	As (33) 947; 77	Se (34) 941; 195	Br (35) 1140; 325
5 (K₇)	Rb (37) 403; 46.9	Sr (38) 550; -146	In (49) 558; 34	Sn (50) 709; 121	Sb (51) 834; 101	Te (52) 869; 190	I (53) 1008; 295
6 (K₉)	Cs (55) 376; 45.5	Ba (56) 503; -46	Tl (81) 589; 30	Pb (82) 716; 35.2	Bi (83) 703; 101	Po (84) 812; 186	At (85) 930; 270

^a For each atom in parentheses is the atomic number, Z . Along the first column, in parenthesis, are the K_p odd complete graphs.

values that show the strongest similarity with the first ionization energies followed by the K_{pp} -(p-odd) values. The electron affinity stands alone with no evident similarity with any K_{pp} or K_p values. All in all, the K_{pp} values with p odd are good descriptors for these atomic properties.

A QSPR Model of Organic and Inorganic Compounds.

We will now consider the models of two properties: the average molecular polarizability, $\langle\alpha\rangle$, and the lattice enthalpy of metal halides, ΔH_L^θ . For both models we will consider $q = 1$, and p , with p odd (see eq 5), i.e., $p = 1, 3, 5$, and 7 , and with p sequential, i.e., $p = 1, 2, 3$, and 4 . The values of these two properties are collected in Tables 3 and 4 respectively, and they are taken from refs 5 and 6. For the metal halides due to their deceptively simple chemical graph made up of two connected vertices, $\bullet-\bullet$, only the chemical pseudograph has a meaning, and this means that only the valence connectivity basis indices, $\{D^v, {}^0\chi^v, {}^1\chi^v\}$, and the pseudoindeces, ψ_{LE} , are meaningful. Further, the total indices equal the first-order basis indices: ${}^1\chi^v \equiv \chi_t^v$ and ${}^1\psi_{LE} \equiv {}^T\psi_{LE}$. These characteristics, and the fact that these compounds cover a wide range of ionic character,^{15–17} render the model of the property of these compounds, with different K_p or K_{pp} representations, worthy to be undertaken.

The Average Molecular Polarizabilities of Fifty-Four Compounds, $\langle\alpha\rangle$ $q = 1$, K_p Description, p Odd, K_{pp} -(p-Odd). This is the overall best description. The best single-, two-, three and four-basis-index descriptors for this property are ($s_0 = 1.15$)

n	$\{\beta\}$	Q	F	r	s_R	\mathbf{u}
54	$\{{}^0\chi^v\}$	0.829	545	0.955	1	(23, 0.9)
54	$\{{}^0\chi^v, {}^0\psi_1\}$	1.379	755	0.984	1.6	(19, 6.1, 0.9)
54	$\{{}^0\chi^v, {}^1\chi^v, {}^S\psi_1\}$	2.012	1071	0.992	2.3	(31, 12, 7.2, 0.1)
54	$\{{}^0\chi^v, {}^1\chi^v, {}^S\psi_1, {}^T\psi_1\}$	2.092	868	0.993	2.4	(29, 7.9, 7.7, 2.2, 1.4)

The best higher-order term contributed by a Z term plus two dual basis indices is the following one

$$Z' = [Z + 0.007 \cdot {}^0\chi_d^v - 0.0004 \cdot {}^1\psi_{ld}]$$

$$Q = 1.704, F = 2304, n = 54, r = 0.989, s_R = 3.5,$$

$$\mathbf{u} = (48, 2.7), \mathbf{C} = (0.62672, -0.50945)$$

Here, $Z = [X + 0.4Y - 0.1({}^0\chi)^{0.1} - 0.9({}^1\psi_E)^{1.5}]$: $F = 1774$, $r = 0.986$, $s_R = 3.1$; $X = [3 \cdot {}^0\chi^v + {}^1\chi]$: $F = 1587$, $r = 0.984$, $s_R = 1.6$, and $Y = [{}^0\psi_1 + 3.1{}^T\psi_1]^{1.2}$: $F = 343$, $r = 0.932$, $s_R = 1.4$. Two things can be noticed about the Z' term: (i) some of its statistics are much better than the statistics of the four-index LCBI, and (ii) with its correlation vectors, \mathbf{C} , the calculated $\langle\alpha(\mathbf{C})\rangle$ values of Table 1 have been obtained. Actually, fourteen $\langle\alpha\rangle$ values (25% of the values)

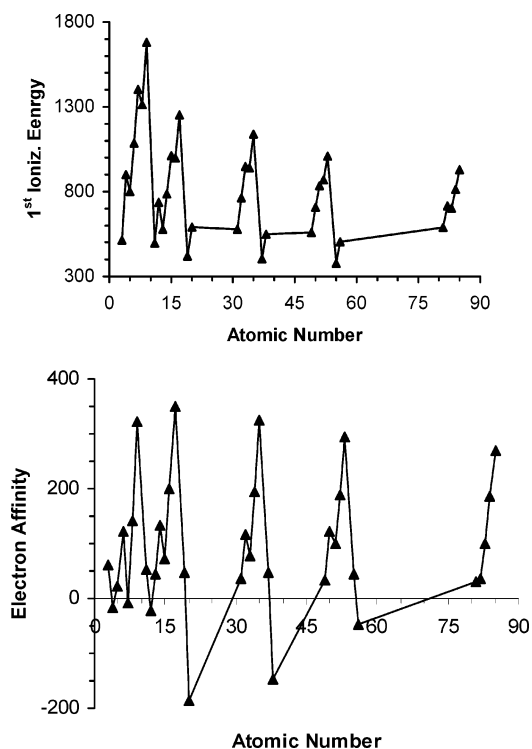


Figure 7. Top – the first ionization energy of atoms 1A–7A vs their atomic number and bottom – the electronic affinity of atoms 1A–7A vs their atomic number.

have been obtained with quantum mechanical methods (denoted with an asterisk in Table 1).¹⁸ These values have been excluded from the model and the remnants forty values have been modeled again with the Z' descriptor, and the statistics now are as follows: $F = 2030$, $n = 40$, $r = 0.991$, $s_R = 2.0$, $\mathbf{u} = (45, 2.8)$, $\mathbf{C} = (0.62734, -0.57659)$. The fourteen left-out values are shown in parentheses in Table 1. The agreement with the corresponding “quantum” values is noteworthy.

$q = p$, K_{pp} Description, p Odd, K_{pp} -(p-Odd). The best Q and F descriptions here are ($s_0 = 1.15$)

$$\{{}^0\chi^v\}:$$

$$Q = 0.422, F = 141, r = 0.855, s_R = 0.57, n = 54$$

$$\{{}^0\chi, \chi_t^v, {}^S\psi_1, {}^T\psi_E\}:$$

$$Q = 0.526, F = 55, r = 0.904, s_R = 0.67, n = 54$$

Both are rather poor descriptions. The basis indices, which are obtained with this algorithm, give rise to no interesting terms.

Table 3. Experimental, $\langle\alpha\rangle$, and Computed, $\langle\alpha(C)\rangle$, Molecular Polarizabilities^a

compound	$\langle\alpha\rangle$	$\langle\alpha(C)\rangle$	compound	$\langle\alpha\rangle$	$\langle\alpha(C)\rangle$
ethane	4.48	4.71	acetaldehyde*	4.59	4.68 (4.62)
propane	6.38	6.46	acetone	6.39	6.94
neopentane	10.20	10.42	F-methane	2.62	3.15
cyclopropane	5.50	5.60	TriF-methane	2.81	4.41
cyclopentane	9.15	9.17	TetraF-methane	2.92	3.44
cyclohexane	11.00	11.00	Cl-methane	4.55	4.71
ethylene	4.12	3.34	DiCl-methane	6.82	6.46
propene*	6.26	5.43 (5.37)	TriCl-methane	8.53	8.40
2Mepropene*	8.29	7.53 (7.47)	TetraCl-methane	10.51	10.42
trans-2-butene*	8.49	7.60 (7.54)	Br-methane	5.61	6.30
cyclohexene*	10.70	10.36 (10.30)	DiBr-methane*	8.68	9.62 (9.56)
butadiene	7.87	6.31	TriBr-methane	11.84	13.09
benzene	9.92	9.13	I-methane	7.59	7.80
toluene*	12.30	11.17 (11.12)	DiI-methane*	12.90	12.63 (12.58)
HexaMebenzene	22.63	22.47	TriI-methane	18.04	17.62
acetylene	3.50	2.71	CH ₂ =CCl ₂	7.83	7.53
propyne	4.68	4.91	cis-CHCl=CHCl	7.78	7.60
C(CCH) ₄	12.19	12.12	disilane*	11.10	11.63 (11.58)
allene	5.00	5.57	formamide*	4.08	3.70 (3.64)
methanol	3.32	3.34	acetamide*	5.67	5.89 (5.83)
ethanol	5.11	5.10	acetonitrile	4.48	4.59
2-propanol*	6.97	7.11 (7.05)	propionitrile*	6.24	6.45 (6.39)
cyclohexanol*	11.56	11.67 (11.61)	Tert-Bucyanide	9.59	10.55
dimethyl ether	5.24	5.37	benzylcyanide	11.97	11.88
p-dioxane	8.60	9.50	TriCl-acetonitrile	10.42	10.55
methylamine	3.59	3.68	pyridine	9.92	9.90
formaldehyde	2.45	2.60	thiophene	9.00	8.67

^aIn parentheses are the values that have been externally modeled (compounds with an asterisk).**Table 4.** Lattice Enthalpies for Twenty Metal Halides, ΔH_L° (kJ mol⁻¹), at 298.15 K, the Calculated Values, $\Delta H_L^\circ(C)$, and the Corresponding Residual, $|\Delta\%|$ ^a

MeX	ΔH_L°	$\Delta H_L^\circ(C)$	$ \Delta\% $	$\Delta H_L^\circ(6\text{-loo})$
LiF*	1037	1034	0.3	1031
NaF	926	902	2.6	
KF	821	825	0.5	
RbF	789	781	1.0	
CsF*	750	752	0.2	752
LiCl	852	886	4.0	
NaCl	786	787	0.2	
KCl	717	732	2.1	
RbCl*	695	702	1.0	702
CsCl	678	681	0.5	
LiBr	815	805	1.2	
NaBr	752	749	0.4	
KBr*	689	683	0.9	683
RbBr	668	660	1.3	
CsBr	654	644	1.5	
LiI	761	763	0.3	
NaI*	705	692	1.7	693
KI	649	656	1.2	
RbI	632	637	0.8	
CsI *	620	624	0.7	625

^a The last column shows the ΔH_L° calculated values for the six left-out compounds, i.e., every fourth, plus the last compound (compounds with an asterisk in the first column).

$q = 1$, K_p Description, p Sequential, K_p -(p-Seq). Here the model gets even worse. In fact the best Q and F descriptors are ($s_0 = 1.15$)

$$\{\chi^v\}: Q = 0.318, F = 80, r = 0.779, s_R = 0.47, n = 54$$

$$\{\chi, \chi^v, {}^1\chi^v, {}^s\psi_I\}: Q = 0.378, F = 28, r = 0.835, \\ s_R = 0.52, n = 54$$

Even here no good terms could be found.

$q = p$, K_{pp} Description, p Sequential, K_{pp} -(p-Seq). Here the model improves over the two previous ones but does not compete with the K_p -(p-Odd) model. The best Q and F descriptors here are ($s_0 = 1.15$):

$$\{\chi^v\}: \\ Q = 0.640, F = 325, r = 0.928, s_R = 0.79, n = 54$$

$$\{\chi^v, \chi_I^v, {}^s\psi_I, {}^T\psi_E\}: Q = 0.790, F = 129, \\ r = 0.954, s_R = 0.95, n = 54$$

Note how throughout these three last descriptions $s_R < 1$. Finally, even here the optimal X and Y terms are worse than the corresponding K_{pp} -p-odd terms, and no interesting Z term could be detected.

Lattice Enthalpy of Twenty Metal Halides, ΔH_L° . $q = 1$, K_p Description, p Odd, K_p -(p-Odd). Even if this is not the optimal description for this property, it shows nevertheless striking features, especially at the level of the two-index LCBI. The best single- and two-basis index descriptors are

$$\{\psi_E\}: \\ Q = 0.023, F = 108, r = 0.926, s_0 = 40.6, n = 20$$

$$\{\chi^v, {}^1\psi_E\}: Q = 0.042, F = 183, r = 0.978, \\ s_R = 1.8, n = 20, \mathbf{u} = (8.1, 11, 31)$$

Actually the two-basis-index combination is the optimal combination for this property. We will not go into the details of the higher-order terms, X , Y , Z and Z' , as the next description shows better terms, even if the difference is not dramatic.

$q = p$, K_{pp} Description, p Odd, K_{pp} -(p-Odd). The best single-basis-index and the higher-order terms show, here, the

optimal statistics. The best single- and two-basis index descriptors, here, are ($s_0 = 40.6$)

$$\{^0\psi_1\}: Q = 0.028, F = 159, r = 0.948, s_R = 1.2, n = 20$$

$$\{^0\psi_1, ^1\psi_1\}: Q = 0.041, F = 173, r = 0.976, \\ s_R = 1.7, n = 20, \mathbf{u} = (8.2, 4.4, 16)$$

The two-basis index combination even if not as good as the previous one is nevertheless rather similar to it. All terms improve in ≈ 1.2 times relative to the terms of the previous case. The following Z' term, contributed by a Z term plus a dual index, could be detected

$$Z' = [Z^{0.9} + 0.0001(^1\chi^v_s)]^{0.2}$$

$$Q = 0.081, F = 1374, r = 0.994, s_R = 3.3, n = 20, \\ \mathbf{u} = (37, 37), \mathbf{C} = (258391, -252589)$$

where

$$Z = (0.6X + 0.1Y^3)^{0.01}: F = 1087, r = 0.992, s_R = 3.0$$

$$X = (D^v)^{0.4}/[0.9D^v + 50(^0\chi^v)^{1.1}]^{0.9}: F = 1035, \\ r = 0.991, s_R = 2.9$$

$$Y = (^0\psi_1)^{0.01}: F = 206, r = 0.959, s_R = 1.3$$

The ΔH_L^\varnothing values, calculated by the aid of this Z' term, ΔH_L^\varnothing , and the corresponding modulus percent residual, $|\Delta\%-(ppo)| = |\Delta H_L^\varnothing - \Delta H_L^\varnothing(K_p)|100/\Delta H_L^\varnothing$, are shown in the last two columns of Table 4. Leaving-out the six compounds (ca. 25% of the compounds) with the asterisk in Table 4 the values of the statistical parameters for Z' become the following: $F = 487, n = 14, r = 0.989, s_R = 2.8, \mathbf{u} = (22, 22), \mathbf{C} = (256613, -25085)$. The calculated ΔH_L^\varnothing values for these six compounds are shown throughout the last column of Table 4. The quality of these values need not be underlined.

$q = 1, K_p$ Description, p Sequential, K_p -(p -Seq). This nonoptimal description for this property shows nevertheless interesting features at the level of the two-index LCBI. The best single- and two-basis index descriptors, the same as those of the previous case, are ($s_0 = 40.6$)

$$\{^0\psi_1\}: Q = 0.023, F = 109, r = 0.927, s_R = 1.0, n = 20$$

$$\{^0\psi_1, ^1\psi_1\}: Q = 0.040, F = 163, r = 0.974, \\ s_R = 1.6, n = 20, \mathbf{u} = (8.6, 5.6, 16)$$

Concerning the higher-order terms, only the Z term comes rather close to the Z term of the previous case, while no optimal Z' could be found here

$$Z = (0.6X + 0.4Y^6)^{0.01}: F = 1052, r = 0.992, s_R = 2.9$$

where

$$X = (D^v)^{0.4}/[D^v + 25(^0\chi^v)^{1.1}]^{0.7}: F = 932, \\ r = 0.990, s_R = 2.7$$

$$Y = (^0\psi_1)^{0.01}: F = 139, r = 0.941, s_R = 1.1$$

$q = p, K_{pp}$ Description, p Sequential, K_{pp} -(p -Seq). Things here improve a lot for the single-basis-index description, which has practically the same model quality of the K_{pp} -(p -odd) case. The two-basis-index combination is deceiving ($s_0 = 40.6$).

$$\{^0\psi_1\}: Q = 0.028, F = 159, r = 0.948, s_R = 1.2, n = 20$$

$$\{^0\psi_1, ^1\psi_1\}: Q = 0.030, F = 91, r = 0.9564, \\ s_R = 1.3, n = 20, \mathbf{u} = (5.3, 1.7, 12)$$

Concerning the higher-order terms, only an interesting Y term could be detected, the other terms are, instead, deceiving

$$Y = (^0\psi_1)^{0.01}: F = 203, r = 0.958, s_R = 1.3$$

CONCLUSIONS

The proposed complete graph model to encode the core electrons for atoms with the principal number $n \geq 2$ offers an interesting possibility to complete the “graph” character of the molecular connectivity theory. Up to date, this theory for atoms with $n > 2$ had to choose two types of δ^v numbers, which were, mainly, based on quantum concepts. The new valence delta number, δ^v , defined with algorithm (5) maintains this double typology for this number, with $q = 1$ and $q = p$, allowing for a new duplicity, i.e. either p odd or p sequential. This twice-double character is actually rather advantageous, as it allows to avoid degeneracy in δ^v , which can arise with some atoms or groups of atoms like, i.e., Cl, Br and $-\text{CH}_3$. Further, it offers the possibility to optimize the encoding of the core electrons for specific classes of compounds.

The atomic properties such as the electronegativity and the curvature are preferentially described by $\delta^v(K_{pp})$ with p odd values, further, for p odd we have a better resolution between the $\delta^v(K_{pp})$ and the $\delta^v(K_p)$ values. For p sequential the resolution among the two set of values lowers. Instead, the $\delta^v(K_{pp})$ values with p sequential are more apt to describe the first ionization energy, while none of the δ^v values shows a good description for the electron affinity.

Up to now, QSPR studies seem to prefer p odd, even if no preference is shown either for $p = 1$ or for $p = q$, as it is exemplified by the present study of the average molecular polarizabilities of organic derivatives and of the lattice enthalpies of metal halides. Work in progress seems to show a preference for the K_{pp} -(p -odd) representation of the core electrons.¹⁹

REFERENCES AND NOTES

- (1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Wiley: New York, 1986.
- (2) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological state*; Academic Press: New York, 1999.
- (3) Pogliani, L. From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors. *Chem. Rev.* **2000**, *100*, 3827–3858.
- (4) Pogliani, L. Algorithmically compressed data and the topological conjecture for the inner-core electrons. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1028–1042.

- (5) Pogliani, L. Modeling with dual indices and complete graphs. The heterogeneous description of the dipole moments and polarizabilities. *New J. Chem.* **2003**, 27, 919–927.
- (6) Pogliani, L. The complete graph conjecture for the inner-core electrons. The homogeneous index case. *J. Comput. Chem.* **2003**, 24, 1097–1109.
- (7) Harary, F.; Read, R. *Proc. Graphs and Combinatorics Conference*, George Washington University; Springer: New York, 1977. Cited in: Barrow, J. D. *The Book of Nothing*; Vintage Books: New York, 2000, p 155.
- (8) Rosen, K. H. *Discrete Mathematics and its Applications*; McGraw-Hill: New York, 1995.
- (9) Pogliani, L. Modeling with molecular pseudoconnectivity descriptors. A useful extension of the intrinsic I-State concept. *J. Phys. Chem. A* **2000**, 104, 9029–9045.
- (10) Pogliani, L. How far are molecular connectivity descriptors from IS molecular pseudoconnectivity descriptors? *J. Chem. Inf. Comput. Sci.* **2001**, 41, 836–847.
- (11) Pogliani, L. Topics in molecular modeling: dual indices, quality of modeling and missing information, truncation. *J. Mol. Struct. (THEOCHEM)* **2002**, 581, 87–109.
- (12) Pogliani, L. Mixed Higher-Order Connectivity-Pseudoconnectivity Terms, *Topology in Chemistry*; Rouvray, D. H., Bruce King, R., Eds.; Horwood: Chichester, 2002; pp 208–248.
- (13) Chang, R. *Chemistry*; McGraw-Hill: New York, 1994.
- (14) Atkins, P. W. *Physical Chemistry*; Oxford: Oxford, 1990.
- (15) Matta, C. F.; Gillespie, R. J. Understanding and interpreting molecular electron density distributions. *J. Chem. Ed.* **2002**, 79, 1141–1152.
- (16) Gillespie, R. J. Electron densities, atomic charges, and ionic, covalent and polar bonds. *J. Chem. Ed.* **2001**, 78, 1688–1692.
- (17) Laing, M. A tetrahedron of bonding. *Ed. Chem.* **1993**, 30, 160–163.
- (18) Ma, B.; Lii, J.-H.; Allinger, N. L. Molecular polarizabilities and induced dipole moments in molecular mechanics. *J. Comput. Chem.* **2000**, 21, 813–825.
- (19) Pogliani, L. work in progress.

CI0341262