

## A NEW APPROACH FOR DEVISING LOCAL GRAPH INVARIANTS: DERIVED TOPOLOGICAL INDICES WITH LOW DEGENERACY AND GOOD CORRELATION ABILITY

P.A. FILIP and T.-S. BALABAN

*Centre of Organic Chemistry, ICECHIM, Bucharest, Roumania*

and

A.T. BALABAN\*

*Polytechnic Institute, Organic Chemistry Department, Splaiul Independentei 313,  
76206 Bucharest, Roumania*

Received 26 May 1986

### Abstract

A new approach is presented for obtaining graph invariants which have very high discriminating ability for different vertices within a graph. These invariants are obtained as the solution set (*local invariant set, LOIS*) of a system of linear equations  $Q \cdot X = R$ , where  $Q$  is a topological matrix derived from the adjacency matrix of the graph, and  $R$  is a column vector which encodes either a topological property (vertex degree, number of vertices in the graph, distance sum) or a chemical property (atomic number). Twenty examples of LOISs are given and their degeneracy and ordering ability of vertices is discussed. Interestingly, in some cases the ordering of vertices obtained by means of these invariants parallels closely the ordering from an entirely different procedure based on Hierarchically Ordered Extended Connectivities which was recently reported. New topological indices are easily constructed from LOISs. Excellent correlations are obtained for the boiling points and vaporization enthalpies of alkanes *versus* the topological index representing the sum of local vertex invariants. Less spectacular correlations with NMR chemical shifts, liquid phase density, partial molal volumes, motor octane numbers of alkanes or cavity surface areas of alcohols emphasize, however, the potential of this approach, which remains to be developed in the near future.

### 1. Introduction and graph theoretical definitions

The aim of this report is to present a new way for obtaining local graph invariants, i.e. a number associated with each *vertex* (point or node) of a graph such that (i) irrespective of any arbitrary numbering of the vertices, a given vertex is always assigned the same number, (ii) topologically equivalent vertices are assigned equal numbers, and (iii) the same invariant should be obtained for a corresponding vertex

\*Corresponding author.

of any isomorphic graph with the given graph. Such a new approach should be of interest both to graph theorists and to chemists. We shall follow up some chemical consequences of the new graph invariants by devising new topological indices and by testing their degeneracy (how many nonisomorphic graphs have the same topological index) and correlational ability.

In molecular graphs, vertices represent atoms and *edges* (lines) represent covalent bonds. Usually, in order to simplify molecular graphs, hydrogen atoms are overlooked. Accordingly, we shall deal only with *hydrogen-depleted graphs* wherein vertices symbolize non-hydrogen atoms. The number of lines meeting at a vertex is called the *degree* of that vertex.

Graphs that can be written (embedded in a plane) so that no edges cross are called *planar graphs*. Acyclic graphs are called *trees*; all trees are planar graphs. We shall ignore disconnected graphs or directed graphs.

The smallest number of edges which must be traversed in going from one vertex to another is called the *distance* between these two vertices; actually, this is a topological distance and not a geometric one. Two vertices at distance one are called *adjacent*. If two adjacent vertices have more than one edge joining them, the bond between the two vertices is a multiple (double, triple, etc.) bond, and the whole graph becomes a *multigraph*.

The adjacencies characterize a graph up to isomorphism, i.e. we can completely reconstruct a graph if we know its adjacencies. A compact way to present them is by means of the adjacency matrix  $A$ . This is a symmetrical (square) matrix with as many rows and columns as there are vertices. Entries  $a_{ij} = a_{ji}$  are *zero* if vertices  $i$  and  $j$  are non-adjacent, *one* if they are adjacent (and for multigraphs the entry may be higher than one, expressing the multiplicity of the bond between vertices  $i$  and  $j$ ). Unless we deal with loop graphs, where a vertex may be bonded with itself, the main diagonal of  $A$  is composed of zeroes.

A more elaborate matrix is the distance matrix  $D$ . It has the same order as the adjacency matrix but the entries  $d_{ij}$  are distances between vertices  $i$  and  $j$ ; therefore,  $D$  has zeroes only on the main diagonal.

Other definitions will be clarified as we need them. For further reading, a selected bibliography is provided [1–5].

## 2. Devising topological indices

In order to devise superior topological indices (TIs), we concerned ourselves initially with the mechanism of constructing a TI. It appears that for most TIs, such a mechanism must have two stages, both being significant for the degeneracy and the correlation abilities of the desired TI. The first stage consists of finding *local vertex invariants* (LOVIs) for each vertex. In the second stage, these invariants are operated upon to produce a number which represents the TI. We shall denote the above stages

as the *assignment* and the *operational* stages, respectively. It is instructive to list a few characteristics of these two stages and to distinguish them in the construction of known TIs.

The assignment stage must encode the topological information of the structure whose TI is desired. The assignment stage can be either purely *topological* if heteroatoms are not distinguished from carbon atoms, or *chemical* if to heteroatoms are assigned distinct numbers from carbon atoms, even when these are topologically equivalent.

Table 1 presents some known TIs for which the assignment and operational stages can be distinguished. The LOVIs on which these TIs are based are exemplified for isopentane (*I*). The result of the assignment stage is a set of LOVIs  $x_i$  (number  $x$  is attached to vertex  $i$ ). We shall denote the set  $X = \{x_i\}$  as the *local invariant set* (LOIS) of the graph (molecule). For convenience, when comparing LOISs of different graphs, we shall assume that their constituent LOVIs form ordered sets. An ideal assignment stage will lead to distinct LOVIs for nonequivalent vertices in any given graph and to different LOISs for nonisomorphic graphs.

The operational stage "operates" on the LOVIs by means of a mathematical formula which may also encompass topological information. Examples of operational stages which do not include the topology of the structure are simple addition of LOVIs, addition of squares of LOVIs, etc., i.e. all vertices are treated identically by the mathematical formula. On the other hand, operations on certain classes of LOVIs, e.g. the geometrical means or the reciprocal geometrical means of LOVIs for any pair of adjacent vertices, are operations which take into account the topology of the examined structure. The result of the operational stage is a single number which represents the final TI. Some operational stages in the construction of known TIs are also detailed in table 1.

An equivalent construction of TIs may result if one assigns to each edge  $i$  a *local edge invariant*  $y_i$ , thus forming a LOIS  $Y = \{y_i\}$ , and operating on  $Y$ . One can calculate, for example, TIs  $w$ ,  $\chi$ , and  $J$  for *I* by first assigning to each bond the  $y_i$  values shown in fig. 1, and then adding these values; in these examples, the same operational formula (addition) was applied for all three TIs, although this was not the case when LOVIs were used (see table 1). Other examples include TIs  $p$  and  $N_2$ , which in fact were defined on the basis of local edge invariants and not on LOVIs.

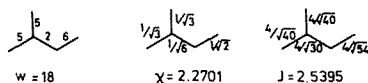









Fig. 1. Local edge invariants for isopentane *I*.

It is thus possible to change the topological content of the assignment or operational stages when using local edge invariants instead of LOVIs. In the following discussion, we shall be concerned only with LOVIs.

Table 1  
Some topological indices for which the assignment and operational stages can be distinguished

Topological Index (TI)	Ref.	Symbol	Assignment stage		Operational stage	
			Local Vertex Invariant (LOVI), $x_i$	Example: $I$	Computing formula <sup>a</sup>	TI for $I$
Gordon-Scantlebury	[6]	$N_2$	Number of second neighbours <sup>b</sup>		$\frac{1}{2} \sum x_i$	4
Polarity no.	[7]	$p$	Number of third neighbours <sup>b</sup>		$\frac{1}{2} \sum x_i$	2
Gutman et al.	[8]	$M_1$	Vertex degrees		$\sum x_i^2$	16
Gutman-Trinajstić	[9]	$M_2$	Vertex degrees		$\sum x_i x_j$	14
Randić	[10]	$\chi$	Vertex degrees		$\sum (x_i x_j)^{-1/2}$	2.2701
Wiener	[7]	$w$	Distance sum		$\frac{1}{2} \sum x_i$	18
Balaban	[11, 12]	$J$	Distance sum		$\frac{q}{\mu + 1} \cdot \sum (x_i x_j)^{-1/2}$ <sup>c</sup>	2.5395

<sup>a</sup>The summations are for all vertices  $i$ , or for all edges  $i-j$  in the graph.

<sup>b</sup>Originally defined on the basis of local edge invariants (see text).

<sup>c</sup> $q$  represents the total number of edges  $i-j$ ;  $\mu$  represents the cyclomatic number.

In the construction of other TIs, these two stages can not be distinguished. Some TIs are derived by a more direct approach, e.g. Lovasz's lowest eigenvalue [13], or do not make use of vertex or edge invariants, e.g. Hosoya's index  $Z$  [14], or Bonchev's information indices [15]. For reviews on TIs, refs. [11], [16] and [17] should be consulted.

The degeneracy of an index may be due either to the assignment stage or, accidentally, to the operational stage. *Assignment degeneracy* results if nonequivalent vertices receive identical LOVIs, or if nonisomorphic graphs have the same ordered LOISs. *Operational degeneracy* is encountered seldom and leads to the same TI for nonisomorphic graphs which have no assignment degeneracy. Simple operational stages on weakly differentiated LOVIs will certainly lead to highly degenerate TIs. Usually one has to compromise between the complexity of the two stages and the computing time in order to obtain low degeneracy for the final index.

In the present paper we present a new approach for the assignment stage. We need to construct a LOIS as close as possible to the ideal, i.e. with no assignment degeneracy, but whose construction must remain simple and unequivocal. This approach may prove to lead to the lowest assignment degeneracy attained so far. The resulting LOISs can be employed either in various procedures for ordering structures or in operational stages to yield new TIs. Examples will be given both for the structure ordering and for the correlation abilities of some new TIs. Some of these correlations give better results than previous correlations using other TIs.

### 3. New graph invariants. Definitions, examples and assignment degeneracy

Our approach for obtaining a local invariant set (LOIS)  $X$  of local vertex invariants (LOVIs)  $x_i$  consists of solving a linear system of equations:

$$Q \cdot X = R, \quad (1)$$

where  $Q$  is a matrix derived from the adjacency matrix,  $R$  is a column vector and  $X$  is the column vector of LOVIs. Thus, for a graph with  $N$  vertices, the LOVIs  $x_i$  ( $i = 1, 2, \dots, N$ ) are obtained by solving a system of  $N$  linear equations. The numerical values of the LOVIs and hence their properties and utility will depend strongly on the chosen matrices  $Q$  and  $R$ . By suitably choosing these matrices, a large variety of LOISs can be obtained. We now describe the way in which matrices  $Q$  and  $R$  can be constructed.

Matrix  $Q$  is derived from the adjacency matrix  $A$ , or different powers of the adjacency matrix ( $A^2, A^3, \dots, A^N$ ) by replacing the diagonal elements  $a_{ii}$  (associated with vertex  $i$ ) with a nonzero parameter  $p_i$ . This parameter  $p_i$  describes a certain property  $P$  of vertex  $i$ . In turn, this property can be either *topological*, e.g. the vertex

degree  $V = \{v_i\}$ , the distance sum  $S = \{s_i\}$ , the total number of vertices in the graph  $N$ , or *chemical*, e.g. the atomic number  $Z = \{z_i\}$ , ionization potential, electronegativity, etc. The column matrix  $R$  consists of parameters  $r_i$  which may or may not be identical with  $p_i$ , thus describing the same or another property of vertex  $i$ . If  $P$  is the property attributed to the diagonal terms of matrix  $A^\alpha$  ( $\alpha = 1, \dots, N$ ) and  $R$  is the property described by the column matrix, the system (1) can be written as:

$$(A^\alpha + P \cdot \mathbf{1}) \cdot X = R, \quad (2)$$

where  $\mathbf{1}$  is the unity matrix and  $\alpha$  is the power to which the adjacency matrix is raised. We shall operate only with the adjacency matrix ( $\alpha = 1$ ) or with the distance matrix  $D$ . A simpler way to denote system (2) is APR (or DPR if the distance matrix instead of the adjacency matrix is employed).

Figure 2 illustrates the procedure for obtaining a LOIS for the three isomers 1–3 of pentane using the adjacency matrix  $A$ , the atomic numbers  $Z$  as property  $P$ , and the vertex degree  $V$  as property  $R$ . We shall denote this combination of parameters as describing the equation system-AZV. The procedure outlined in fig. 2 has the following steps: (i) the vertices are arbitrarily numbered; (ii) the adjacency matrix is constructed; (iii) the diagonal zeroes are replaced by the atomic number  $z_i$  of the corresponding vertex, i.e. six (all atoms being carbon atoms); (iv) the column matrix  $V$  is constructed from the vertex degrees of the appropriate vertices; and (v) the system AZV is solved numerically. The LOVIs  $x_1 - x_5$  thus obtained are indicated in the last column of fig. 2. Another example of LOVIs obtained analogously by solving the system-AZV is presented in fig. 3 for all isomers of pentane (1–3), butylamine (4–12) and butylborane (13–21). A brief inspection of the LOVIs reveals that the heteroatom (nitrogen with  $z = 7$  or boron with  $z = 5$ ) influences markedly the value of the LOVIs for the respective and adjacent vertices, while remote vertices remain unaffected. It should also be noted that the AZV-LOVI increases when either the atomic number of that vertex decreases, or the atomic numbers of adjacent vertices increase.

Instead of solving numerically the system AZV, it is possible in special cases to obtain analytical expressions for LOVIs; simple linear systems are obtained, for example, for the AZV-LOVIs of linear alkanes, as will be shown in a separate paper.

As general trends for these AZV-LOVIs, the following characteristics may be distinguished: (i) all AZV-LOVIs are positive and their numerical range is approximately between 0.1 and 1; (ii) AZV-LOVIs increase from the end of the linear alkane towards its center; (iii) this increase is alternant (not monotonic) and asymptotic; (iv) as the number  $N$  of vertices increases in the linear alkane, the LOVIs increase asymptotically.

If system (2) is solved using other properties instead of the atomic number  $Z$  for matrix  $P$  and the vertex degrees  $V$  for matrix  $R$ , different LOISs will be obtained. In table 2, twenty LOISs are listed together with the general trends of their LOVIs for the case of linear alkanes. The first entry, AZV, is the one presented above.

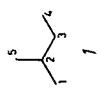
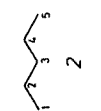
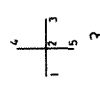
Step: i	ii	iii	iv	v	X
Pentane isomer	A	A + Z · I	V	System (2)	X
	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 6 & 1 & 0 & 0 & 0 \\ 1 & 6 & 1 & 0 & 1 \\ 0 & 1 & 6 & 1 & 0 \\ 0 & 0 & 1 & 6 & 0 \\ 0 & 1 & 0 & 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 3 \\ 2 \\ 1 \\ 1 \end{pmatrix}$	$\begin{cases} 6x_1 + x_2 \\ x_1 + 6x_2 + x_3 + x_5 \\ x_2 + 6x_3 + x_4 \\ x_3 + 6x_4 + x_5 \\ x_2 \\ x_2 \end{cases}$	$\begin{cases} x_1 = 0.09532 \\ x_2 = 0.42808 \\ x_3 = 0.24090 \\ x_4 = 0.12652 \\ x_5 = 0.09532 \end{cases}$
	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 6 & 1 & 0 & 0 & 0 \\ 1 & 6 & 1 & 0 & 0 \\ 0 & 1 & 6 & 1 & 0 \\ 0 & 0 & 1 & 6 & 1 \\ 0 & 0 & 0 & 1 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 1 \end{pmatrix}$	$\begin{cases} 6x_1 + x_2 \\ x_1 + 6x_2 + x_3 \\ x_2 + 6x_3 + x_4 \\ x_3 + 6x_4 + x_5 \\ x_4 + 6x_5 \end{cases}$	$\begin{cases} x_1 = 0.12121 \\ x_2 = 0.27273 \\ x_3 = 0.24242 \\ x_4 = 0.27273 \\ x_5 = 0.12121 \end{cases}$
	$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 6 & 1 & 0 & 0 & 0 \\ 1 & 6 & 1 & 1 & 1 \\ 0 & 1 & 6 & 0 & 0 \\ 0 & 1 & 0 & 6 & 0 \\ 0 & 1 & 0 & 0 & 6 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 4 \\ 1 \\ 1 \\ 1 \end{pmatrix}$	$\begin{cases} 6x_1 + x_2 \\ x_1 + 6x_2 + x_3 + x_4 + x_5 \\ x_2 + 6x_3 \\ x_2 + 6x_4 \\ x_2 \end{cases}$	$\begin{cases} x_1 = 0.06250 \\ x_2 = 0.62500 \\ x_3 = 0.06250 \\ x_4 = 0.06250 \\ x_5 = 0.06250 \end{cases}$

Fig. 2. Procedure for obtaining the AZV-LOIS  $X = \{x_1, x_2, x_3, x_4, x_5\}$  for pentane isomers 1 - 3.

Fig. 3.

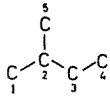
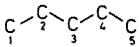
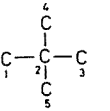
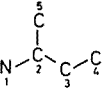
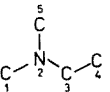
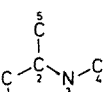
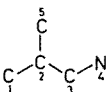
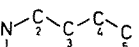
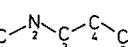
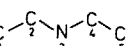
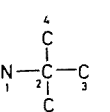
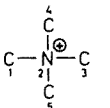
Graph	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1 	0.0953	0.4281	0.2409	0.1265	0.0953
2 	0.1212	0.2727	0.2424	0.2727	0.1212
3 	0.0625	0.6250	0.0625	0.0625	0.0625
4 	0.0814	0.4305	0.2405	0.1266	0.0949
5 	0.1063	0.3622	0.2522	0.1246	0.1063
6 	0.0943	0.4345	0.2047	0.1325	0.0943
7 	0.0954	0.4275	0.2441	0.1080	0.0954
8 	0.1035	0.2758	0.2413	0.2728	0.1212
9 	0.1280	0.2318	0.2494	0.2715	0.1214
10 	0.1202	0.2790	0.2060	0.2790	0.1202
11 	0.0533	0.6267	0.0622	0.0622	0.0622
12 	0.0789	0.5263	0.0789	0.0789	0.0789



Fig. 3 (continued)

	Graph	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
13		0.1151	0.4246	0.2415	0.1264	0.0959
14		0.0794	0.5233	0.2246	0.1292	0.0794
15		0.0968	0.4189	0.2926	0.1179	0.0968
16		0.0952	0.4289	0.2364	0.1527	0.0952
17		0.1463	0.2684	0.2432	0.2726	0.1212
18		0.1115	0.3312	0.2324	0.2744	0.1209
19		0.1227	0.2638	0.2945	0.2638	0.1227
20		0.0755	0.6226	0.0629	0.0629	0.0629
21		0.0385	0.7692	0.0385	0.0385	0.0385

Fig. 3. AZV-LOVIs for hydrogen-depleted graphs of pentane, butylamine and butylborane isomers.

Table 2  
LOVIs for linear alkanes, general trends

LOIS No.	Type	Numerical range	Variation from the endpoints towards the center ( $x_1 \rightarrow x_n$ )	Variation with increasing chain length for any $x_i$
1	AZV	0.1 – 1	alternant increase	increase
2	ASV	0.01 – 0.2	monotonic increase	decrease
3	DSV	-0.02 – 0.12	monotonic increase	increase
4	AZS	2 – 9	monotonic decrease	increase
5	ASZ	0.1 – 1	monotonic increase	increase
6	DN <sup>2</sup> S	0.1 – 0.3	monotonic decrease	increase
7	DN <sup>2</sup> I	0 – 0.09	monotonic increase	decrease
8	AS1	0.02 – 0.1	monotonic increase	decrease
9	DS1	0 – 0.3	monotonic increase	decrease
10	ASN	0.2 – 0.7	monotonic increase	decrease
11	DSN	0.05 – 0.7	monotonic increase	decrease
12	DN <sup>2</sup> N	0.06 – 0.2	monotonic increase	decrease
13	ANS	1 – 4	monotonic increase	increase
14	ANV	0.08 – 0.5	alternant increase	decrease
15	AZN	0.3 – 1.5	alternant increase	increase
16	ANZ	0.5 – 1.7	alternant decrease	decrease
17	AN1	0.1 – 0.3	alternant decrease	decrease
18	DSZ	0.06 – 0.6	monotonic increase	decrease
19	ANN	0.7 – 0.9	alternant decrease	increase
20	DN <sup>2</sup> Z	0.03 – 0.5	monotonic increase	decrease

As a final example, table 3 presents the LOVIs for pentane isomers 1 – 3 in all the twenty combinations defined in table 2.

As can be seen from the above examples, this type of assignment stage which involves solving a system of linear equations leads to a wide range of numerical values for the LOVIs (see table 2), and thus has good chances to yield low assignment degeneracy. Moreover, by construction, one fulfils the condition that equivalent vertices (either topologically or chemically equivalent) will receive identical LOVIs, because these appear in identical equations in the system. This fact is shown by LOVIs  $x_1$  and  $x_5$  for 1,  $x_1$  and  $x_5$  or  $x_2$  and  $x_4$  for 2 and by  $x_1$ ,  $x_3$ ,  $x_4$ , and  $x_5$  for 3, all having equal values, respectively, and have therefore been listed only once in table 3.

When only graph-theoretical properties are included in matrices  $P$  and  $R$ , the resulting LOISs are good candidates for graph-theoretical applications, e.g. for hierarchical ordering of structures.

We have computed the AZV-LOISs (no. 1 in table 2) for all alkanes with up to 12 carbon atoms. No assignment degeneracy was found, i.e. in all cases distinct LOVIs were obtained for nonequivalent atoms. Moreover, all isomers have different (ordered)

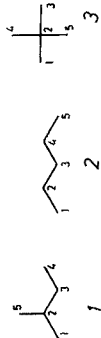


Table 3

LOIs for pentane isomers 1 – 3

LOIS No.	Type	1					2					3				
		$x_1 = x_5$	$x_2$	$x_3$	$x_4$	$x_5$	$x_1 = x_5$	$x_2 = x_4$	$x_3$	$x_4$	$x_5$	$x_1 = x_3 = x_4 = x_5$	$x_2$	$x_3$	$x_4$	$x_5$
1	AZV	0.0953	0.4281	0.2409	0.1265	0.1212	0.2727	0.2424	0.0625	0.6250		0.0625	0.2424	0.2424	0.0625	0.6250
2	ASV	0.0587	0.5304	0.2307	0.0855	0.0761	0.2386	0.2538	0.0000	1.0000		0.0000	0.2538	0.0000	1.0000	
3	DSV	0.0096	0.5673	0.2404	-0.0481	-0.0270	0.2027	0.2838	0.0000	1.0000		0.0000	0.2838	0.0000	1.0000	
4	AZS	1.2860	0.2842	0.7227	1.3796	1.5353	0.7879	0.7374	1.1875	-0.1250		1.1875	0.7374	1.1875	-0.1250	
5	ASZ	0.6519	0.7848	0.7724	0.5808	0.5330	0.6701	0.7767	0.7500*	0.7500*		0.7500*	0.7767	0.7500*	0.7500*	
6	DN <sup>2</sup> S	0.2459	0.1505	0.1834	0.2816	0.3041	0.2077	0.1747	0.2218	0.1245		0.2218	0.1747	0.2218	0.1245	
7	DN <sup>2</sup> I	0.0302	0.0340	0.0327	0.0288	0.0278	0.0317	0.0330	0.0311	0.0350		0.0311	0.0330	0.0311	0.0350	
8	AS1	0.1087	0.1308	0.1287	0.0968	0.0888	0.1117	0.1294	0.1250*	0.1250*		0.1250*	0.1294	0.1250*	0.1250*	
9	DS1	0.0565	0.1454	0.0988	0.0301	0.0304	0.0845	0.1182	0.0625	0.1875		0.0625	0.1182	0.0625	0.1875	
10	ASN	0.5433	0.6540	0.6437	0.4840	0.4442	0.5584	0.6472	0.6250*	0.6250*		0.6250*	0.6472	0.6250*	0.6250*	
11	DSN	0.2825	0.7272	0.4988	0.1502	0.1520	0.4223	0.5912	0.3125	0.9375		0.3125	0.5912	0.3125	0.9375	
12	DN <sup>2</sup> N	0.1508	0.1699	0.1633	0.1437	0.1392	0.1585	0.1651	0.1556	0.1751		0.1556	0.1651	0.1556	0.1751	
13	ANS	1.5579	0.2106	0.8311	1.6338	1.8273	0.8636	0.8546	1.4762	-0.3810		1.4762	0.8546	1.4762	-0.3810	
14	ANV	0.0987	0.5066	0.2695	0.1461	0.1364	0.3182	0.2727	0.0476	0.7619		0.0476	0.2727	0.0476	0.7619	
15	AZN	0.7539	0.4766	0.6326	0.7279	0.7323	0.6061	0.6313	0.7813	0.3125		0.7813	0.6313	0.7813	0.3125	
16	ANZ	1.0816	0.5920	0.8767	1.0247	1.0364	0.8182	0.8727	1.1429	0.2857		1.1429	0.8727	1.1429	0.2857	
17	ANI	0.1803	0.0987	0.1461	0.1708	0.1727	0.1364	0.1455	0.1905	0.0476		0.1905	0.1455	0.1905	0.0476	
18	DSZ	0.3389	0.8726	0.5986	0.1803	0.1824	0.5068	0.7095	0.3750	1.1250		0.3750	0.7095	0.3750	1.1250	
19	ANN	0.9013	0.4934	0.7306	0.8539	0.8636	0.6818	0.7273	0.9524	0.2381		0.9524	0.7273	0.9524	0.2381	
20	DN <sup>2</sup> Z	0.1810	0.2039	0.1960	0.1724	0.1670	0.1901	0.1981	0.1868	0.2101		0.1981	0.1981	0.1868	0.2101	

\* Assignment degeneracy.

LOISs. In a few cases, for two different LOISs, the same LOVI was encountered when less than five significant digits were used for calculations, but on raising the computational precision, such LOVIs became distinct. The remaining LOISs (nos. 2–20) were tested for all alkanes with less than nine carbon atoms, and nonexhaustively for alkanes with up to ten carbon atoms. Only one assignment degeneracy was found, namely for 3 in the case of LOVIs ASZ, AS1 and ASN. Such a triad, denoted AST, differing only in the constant last term ( $T = Z, 1, \text{ or } N$ ) yields practically the same inter- and intramolecular invariant ratios; other such triads are DST,  $DN^2T$  or ANT.

Various cyclic graphs were also tested for assignment degeneracy and only one general example was found in which two classes of nonisomorphic graphs have the same LOISs. These graphs are presented in fig. 4 (22–27) and consist of a ring of  $2k$

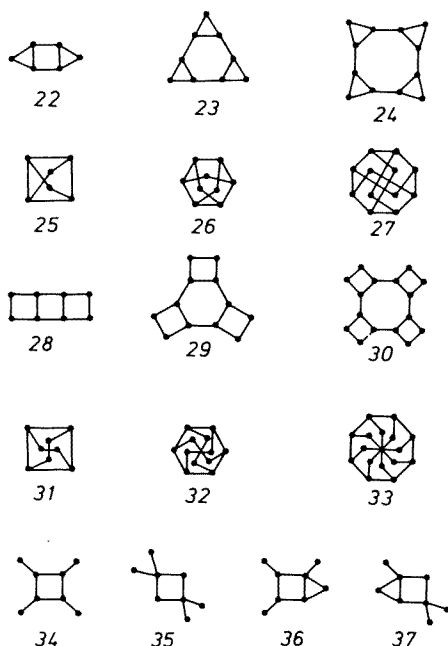


Fig. 4. Cyclic graphs with assignment degeneracy for various LOISs (see text).

vertices linked pairwise by means of  $k$  one-vertex bridges; in one class, each of the  $k$  bridges links two adjacent vertices on the ring (22–24), and in the other class, each bridge links two opposite vertices on the ring (25–27). None of the twenty LOISs differentiated these two classes of cyclic graphs. When the ring vertices are linked by  $k$  bridges consisting of two vertices, as in graphs 28–30 (for adjacent ring vertices) and 31–33 (for opposite ring vertices), the assignment degeneracy between these two classes of graphs is lifted in all cases involving the distance matrix or the distance sums ( $D$  or  $S$ ), but persists otherwise, i.e. for LOISs nos. 1, 14, 15, 16, 17 and 19. This degeneracy also persists if longer bridges (more than two vertices) are used for linking adjacent or opposite ring vertices.

In contrast, graphs 34 and 35 present assignment degeneracy in all cases which *do not* include the adjacency matrix or the vertex degrees ( $A$  or  $V$ ), i.e. for LOISs nos. 6, 7, 9, 11, 12, 18 and 20.

These examples of assignment degeneracy arise from the pairwise identity of systems (2) for the isomeric graphs presented above when these systems involve (i) *both* the adjacency *and* distance matrices (graphs 22 – 27), (ii) *only* the adjacency matrix (graphs 28 – 33), or (iii) *only* the distance matrix (graphs 34 and 35).

For other cyclic graphs, e.g. 36 and 37 which have the same topological index  $J$  [18], most LOISs are nondegenerate even when less than five significant digits are used for computations, but in other cases, especially when the LOVIs are small (e.g. for nos. 7 and 20) and not well distinguished, one has to increase the precision in calculations in order to obtain nondegenerate LOISs.

The LOISs were generated by means of a computer program which, from a standard input (all vertex adjacencies), constructs the topological matrices  $A$  and  $D$ , calculates vertex degrees, distance sums, etc. and, according to the desired LOISs, assembles the appropriate combination for the system (2) of linear equations. This system is then solved with a preimposed precision by a simple and rapid iterative algorithm such as Gauss–Seidel [19] which is convergent for all LOISs, as the diagonal elements in the system matrix  $Q$  are significantly larger than all other elements. Usually ten to fifteen iterations are needed for convergence to the tenth decimal for all LOVIs.

#### 4. New topological indices and operational degeneracy

Using the described assignment stage, various local invariant sets (LOISs) were devised and computed. Using the local vertex invariants (LOVIs)  $x_i$  from these LOISs, we constructed new topological indices (TIs) by means of the operational stages defined by formulae (3)–(7):

$$\text{TI}(3) = \sum x_i \quad (3)$$

$$\text{TI}(4) = \sum x_i^2 \quad (4)$$

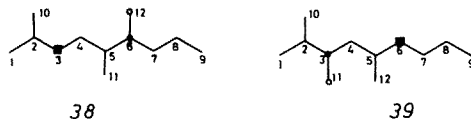
$$\text{TI}(5) = \sum x_i^{1/2} \quad (5)$$

$$\text{TI}(6) = \sum (x_i \cdot x_j)^{-1/2} \quad (6)$$

$$\text{TI}(7) = N \cdot (\prod x_i)^{1/N}. \quad (7)$$

Operations in eqs. (3)–(5) and (7) refer to all vertices  $i$ , and in eq. (6) to all edges  $i-j$ . Topological indices combining the above formulae with LOISs nos. 1–20 were

computed for all alkanes with up to nine carbon atoms, as well as for the structures which were investigated for correlations and are given in the next paragraphs. The AZV-LOIS was used to calculate TIs (3), (6) and (7) for all alkanes with up to twelve carbon atoms, usually six significant digits being enough to discriminate most structures. However, an operational degeneracy was found, namely for the TI defined by the simplest operational formula (3), in the case of the dodecane isomers 38 and 39,



	$x_1 = x_{10}$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
38	0.09478	0.43130	0.22264	0.23285	0.38028	0.38218
39	0.09476	0.40314	0.38218	0.20078	0.41313	0.22264

	$x_7$	$x_8$	$x_9$	$x_{11}$	$x_{12}$
38	0.22364	0.27595	0.12068	0.10329	0.10297
39	0.25102	0.27125	0.12146	0.10297	0.09781

	TI(3)	TI(4)	TI(5)	TI(6)	TI(7)
38	2.665341944*	0.76044809	5.44961	46.50697	2.28833
39	2.665341944*	0.76044817	6.09239	46.50134	2.29103

Fig. 5. LOVIs and TIs for dodecane isomers 38 and 39. Vertices with equal LOVIs are indicated. All numerical values are calculated with ten significant digits. Operational degeneracy is denoted by an asterisk.

depicted together with their AZV-LOVIs in fig. 5. This degeneracy persisted even when the LOVIs were computed to ten significant digits, thus demonstrating that simple operational stages may lead to operational degeneracies. Rather frequent operational degeneracies were encountered for TIs (3), (5), and especially (4), when less than six significant digits were used for calculations in the case of LOISs with small values for LOVIs (nos. 7 and 20).

Earlier, the lowest degeneracy of a simple TI had been found for index  $J$  [18,20], which is nondegenerate for alkanes with up to eleven carbon atoms. The dodecane isomers for which  $J$  presents assignment degeneracy are well differentiated by most LOISs (exceptions being nos. 7 and 20, for which more than six significant

digits are needed to discriminate these isomers). This makes the present approach superior; even when simple operational stages are used, e.g. (3), nondegenerate TIs are obtained.

## 5. Chemical applications

### 5.1. HIERARCHICAL ORDERING OF STRUCTURES

Structures can be ordered hierarchically on the basis of their LOISs. Previous attempts for structure ordering have proven useful in various structure coding and retrieval systems [21]. We have investigated the ordering abilities of LOISs nos. 2 (ASV), 3 (DSV), 4 (AZS), 18 (DSZ) and 20 (DN<sup>2</sup>Z) (cf. table 2). An interesting similarity was found between the LOIS-based vertex ordering and an ordering procedure developed recently [22] and based on Hierarchically Ordered Extended Connectivities (HOC) [23].

LOVIs can be used for ordering in two ways: (i) for *vertex ordering* within a structure by giving priority to the vertex with a larger (or smaller) LOVI, and (ii) for *structure ordering* by giving priority to the structure with the larger (or smaller) previously ordered LOIS, the comparison of strings of numbers being easily dealt with by means of various algorithms.

The vertex ordering as obtained by the HOC algorithm [23] for hexane isomers 40–44 is presented in fig. 6. Exactly the same vertex ordering is obtained for 40–44 using LOIS-DSV in increasing order and LOIS-AZS in decreasing order. LOIS-ASV

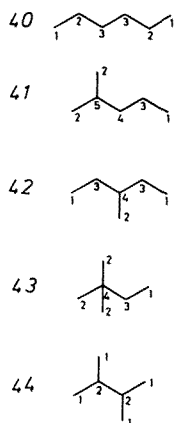


Fig. 6. Vertex ordering of hexane isomers.

used in increasing order of its LOVIs yields practically the same vertex ordering with one exception, namely the first two vertices are in the reverse order for 41–43 than in fig. 6. The other two LOISs (DSZ and DN<sup>2</sup>Z) lead to more numerous inversions

in their vertex ordering. This similar vertex ordering by totally different procedures prompted us to investigate the vertex ordering of all alkanes with nonequivalent vertices with up to thirteen carbon atoms (identity trees) by means of LOISs ASV, DSV and AZS and to compare these with HOC ordering.

A good overall correspondence was found between HOC vertex order and the vertex order obtained with these LOISs, the best fit being obtained with LOIS-DSV. The latter yields the same order as the HOC procedure for alkanes with up to nine vertices and has relatively few inversions for higher alkanes. Interestingly, these inversions arise from the preimposed conventional rules of the HOC algorithm. Without going into details (for which the original HOC papers should be consulted), it seems that the LOIS approach may be superior in discriminating between vertices, no preimposed rules being necessary. On the other hand, the HOC algorithm is solved more rapidly. It might be rewarding to revise the preimposed rules of the HOC procedure in order to obtain the same ordering as by the LOIS approach (or vice versa, to devise a LOIS which gives the same ordering as the HOC procedure).

This similarity points to the fact that graph-theoretical approaches usually have a deeply rooted significance. The graph-theoretical basis of the HMO method is such an example. Why the HOC procedure and the ordering resulting from the LOIS approach are so similar remains to be explained.

## 5.2. CORRELATIONS WITH CHEMICAL SHIFTS

The HOC ordering was found to be similar to the order of  $^1\text{H}$ -NMR chemical shifts of polycyclic aromatic hydrocarbons (PAHs) [22]. Inversions between HOC ordering and the order of chemical shifts were encountered in the bay and the *K* regions of PAHs. As the LOISs nos. 2 (ASV), 3 (DSV) and 4 (AZS) give practically the same ordering as the HOC algorithm, it was expected that these LOISs will correlate at least qualitatively and intramolecularly with the  $^1\text{H}$ -NMR chemical shifts of PAHs. It seemed, therefore, worthwhile to investigate the correlation abilities of LOISs ASV, DSV and AZS with chemical shifts and to see whether these correlations are quantitative and if they can be used for intermolecular comparisons. Apparently, no simple correlation exists between the  $^1\text{H}$ -NMR chemical shifts of the investigated PAHs (anthracene, phenanthrene, triphenylene, chrysene and benz[*a*]anthracene) and LOISs ASV, DSV and AZS. Although a definite correspondence between the LOISs and the chemical shifts exists (i.e. a qualitative correlation, as with HOC ranks), it can not be extended for intermolecular comparisons, especially for LOISs ASV and AZS. Interestingly, the protons which have the largest deviations are the bay-region and *L*-region protons, a fact which might serve as a starting point in studies of carcinogenicity *versus* LOISs of PAHs.

Investigation of  $^{13}\text{C}$ -NMR chemical shifts with LOISs of PAHs and correlations of  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR chemical shifts of acyclic compounds with LOISs are currently under investigation and will be reported separately.



## 5.3. CORRELATIONS OF NEW TOPOLOGICAL INDICES WITH ALKANE BOILING POINTS

The topological index defined for some of the new graph invariants by the simplest operational stage, namely the simple addition of LOVIs according to formula (3), was found to correlate extremely well with the boiling points of alkanes. Twenty-one alkanes (from ethane to *n*-heptane) were considered for correlations with TI (3) calculated by means of all twenty LOISs. Poor correlations resulted when employing LOISs ASV, DSV, DS1, ASN, DSN, DN<sup>2</sup>N, ANZ, AN1, DSZ and ANN. This is in some cases due to the narrow range of variation for TI (3), e.g. for LOIS DN<sup>2</sup>N this range is between 0.77 and 0.8 for all the twenty-one alkanes. In other cases, e.g. when LOISs ASN or DSN are employed, TI (3) yields straight lines separately for the b.p. of each class of alkane isomers.

Good correlations ( $r^2 > 0.95$  and standard deviation  $\bar{s} < 12^\circ$ ) resulted with TI (3) based on the remaining LOISs, the best ( $r^2 = 0.9983$ ,  $\bar{s} = 2.0^\circ$ ,  $n = 21$ ) being obtained with LOIS AZV. Consequently, we expanded the number of alkanes to 50 (for which excellent data are available [24]), and from a regression analysis we obtained the quadratic correlation equation (8) with  $r^2 = 0.9966$  and  $\bar{s} = 2.9^\circ$ :

$$\text{b.p. (}^\circ\text{C)} = -30.23 \cdot \text{TI}(3)^2 + 203.25 \cdot \text{TI}(3) - 142.89. \quad (8)$$

The calculated boiling points by means of eq. (8) are listed in table 4. This represents the best correlation achieved until now by a single TI with the boiling point of alkanes. Previous correlations [24] which yielded comparable results made use of Randić's connectivity indices  ${}^m\chi$ , but only when higher-order ( $m > 3$ ) connectivities were also taken into account in multiparametric correlations.

From table 4 one can see that the largest deviations of the calculated b.p. by eq. (8) from the experimental values are encountered for the symmetrically branched alkanes with at least two side chains, e.g. tetramethylbutane ( $8.2^\circ$ ), 2,4-dimethylpentane ( $6.3^\circ$ ), 3-ethyl-3-methylpentane ( $5.8^\circ$ ), 2,5-dimethylhexane ( $5.2^\circ$ ), 2,2,4,4-tetramethylpentane ( $3.2^\circ$ ), as well as for other branched alkanes, e.g. 2,3,3-trimethylpentane ( $6.8^\circ$ ), 2,4-dimethylhexane ( $5.8^\circ$ ), 2,2,4-trimethylpentane ( $4.7^\circ$ ) and 2,4-dimethylheptane ( $4.5^\circ$ ).

For the same data set of fifty alkanes, other LOISs give a standard deviation of at least  $7^\circ$ , thus making the AZV-LOIS the best candidate for future correlations of boiling points for other substrates.

## 5.4. CORRELATION OF TI(3) BASED ON AZV-LOIS WITH VAPORIZATION ENTHALPIES OF ALKANES

According to Trouton's empirical rule, the vaporization enthalpies ( $\Delta H_{\text{vap}}$ ) of alkanes are dependent on the boiling points. The excellent correlation of TI (3) based on AZV-LOIS with alkane boiling points prompted us to also investigate the correla-

Table 4

Calculated and experimental boiling points ( $^{\circ}\text{C}$ ) and vaporization enthalpies  $\Delta H_{\text{vap}}$  ( $\text{kcal} \cdot \text{mol}^{-1}$ ) for alkanes

No.	Alkane	TI (3)	B.p.		$\Delta H_{\text{vap}}$	
			Calc. <sup>b</sup>	Exp. <sup>a</sup>	Calc. <sup>c</sup>	Exp. <sup>a</sup>
1	$\text{C}_2$	0.2857	-87.3	-88.6	2.760	2.264
2	$\text{C}_3$	0.5294	-43.8	-42.2	3.851	3.965
3	2-Me $\text{C}_3$	0.7273	-11.1	-11.7	4.744	4.799
4	$\text{C}_4$	0.7805	-2.7	-0.5	4.985	5.191
5	2,2-Me $_2\text{C}_3$	0.8750	11.8	9.4	5.415	5.345
6	2-Me $\text{C}_4$	0.9861	28.1	27.8	5.923	6.030
7	$\text{C}_5$	1.0303	34.4	36.1	6.125	6.395
8	2,2-Me $_2\text{C}_4$	1.1421	49.8	49.7	6.640	6.651
9	2,3-Me $_2\text{C}_4$	1.2000	57.5	58.0	6.907	6.985
10	2-Me $\text{C}_5$	1.2346	62.0	60.3	7.067	7.160
11	3-Me $\text{C}_5$	1.2437	63.1	63.3	7.109	7.255
12	$\text{C}_6$	1.2803	67.8	68.7	7.278	7.555
13	2,2,3-Me $_3\text{C}_4$	1.3646	78.2	80.9	7.669	7.669
14	2,2-Me $_2\text{C}_5$	1.3891	81.1	79.2	7.783	7.764
15	3,3-Me $_2\text{C}_5$	1.4079	83.3	86.0	7.871	7.901
16	2,4-Me $_2\text{C}_5$	1.4375	86.8	80.5	8.009	7.872
17	2,3-Me $_2\text{C}_5$	1.4562	89.0	89.8	8.096	8.191
18	2-Me $\text{C}_6$	1.4849	92.3	90.1	8.230	8.325
19	3-Me $\text{C}_6$	1.4924	93.1	91.9	8.265	8.391
20	3-Et $\text{C}_5$	1.5000	94.0	93.5	8.301	8.425
21	$\text{C}_7$	1.5303	97.3	98.4	8.442	8.739
22	2,2,4-Me $_3\text{C}_5$	1.5905	103.9	99.2	8.725	8.402
23	2,2,3,3-Me $_4\text{C}_4$	1.5385	98.3	106.5	8.481	8.410
24	2,2-Me $_2\text{C}_6$	1.6397	109.1	106.8	8.956	8.915
25	2,4-Me $_2\text{C}_6$	1.6955	114.8	109.0	9.218	9.029
26	2,5-Me $_2\text{C}_6$	1.6897	114.2	109.0	9.191	9.051
27	2,2,3-Me $_3\text{C}_5$	1.6194	107.0	109.8	8.860	8.826
28	3,3-Me $_2\text{C}_6$	1.6552	110.7	112.0	9.028	8.973
29	2,3,4-Me $_3\text{C}_5$	1.6673	111.9	113.5	9.085	9.014
30	2,3-Me $_2\text{C}_6$	1.7052	115.8	114.0	9.264	9.272
31	2,3,3-Me $_3\text{C}_5$	1.6290	108.0	114.8	8.905	8.897
32	3-Et-2-Me $\text{C}_5$	1.7111	116.4	115.6	9.292	9.209
33	2-Me $\text{C}_7$	1.7348	118.7	117.7	9.403	9.484
34	3,4-Me $_2\text{C}_6$	1.7127	116.5	117.7	9.299	9.316
35	4-Me $\text{C}_7$	1.7411	119.3	117.7	9.433	9.483
36	3-Me $\text{C}_7$	1.7426	119.5	117.0	9.440	9.521
37	3-Et $\text{C}_6$	1.7489	120.1	118.0	9.470	9.476
38	3-Et-3-Me $\text{C}_5$	1.6723	112.5	118.3	9.109	9.081
39	$\text{C}_8$	1.7803	123.1	125.7	9.618	9.915
40	2,2,4,4-Me $_4\text{C}_5$	1.7419	119.4	122.7	9.437	9.580

Table 4 (continued)

No.	Alkane	TI (3)	B.p.		$\Delta H_{\text{vap}}$	
			Calc. <sup>b</sup>	Exp. <sup>a</sup>	Calc. <sup>c</sup>	Exp. <sup>a</sup>
41	2,4-Me <sub>2</sub> C <sub>7</sub>	1.9442	138.0	133.5	-	-
42	2,6-Me <sub>2</sub> C <sub>7</sub>	1.9393	137.6	135.2	-	-
43	2,5-Me <sub>2</sub> C <sub>7</sub>	1.9474	138.3	136.0	-	-
44	3,3-Me <sub>2</sub> C <sub>7</sub>	1.9056	134.6	137.3	-	-
45	2,3-Me <sub>2</sub> C <sub>7</sub>	1.9553	138.9	140.5	-	-
46	4-EtC <sub>7</sub>	1.9978	142.5	141.2	-	-
47	4-MeC <sub>8</sub>	1.9913	142.0	142.4	-	-
48	2-MeC <sub>8</sub>	1.9848	141.4	142.8	-	-
49	3-MeC <sub>8</sub>	1.9926	142.1	143.5	-	-
50	2,7-Me <sub>2</sub> C <sub>8</sub>	2.1894	157.2	159.6	-	-
51	2,2,3-Me <sub>3</sub> C <sub>6</sub>	1.8686	-	-	10.036	9.871
52	2,2,4-Me <sub>3</sub> C <sub>6</sub>	1.8488	-	-	9.942	9.478
53	2,2,5-Me <sub>3</sub> C <sub>6</sub>	1.8447	-	-	9.923	9.580
54	2,3,5-Me <sub>3</sub> C <sub>6</sub>	1.9085	-	-	10.226	9.910
55	2,2,3,3-Me <sub>4</sub> C <sub>5</sub>	1.8014	-	-	9.718	9.871
56	2,2,3,4-Me <sub>4</sub> C <sub>5</sub>	1.8290	-	-	9.489	9.478
57	2,3,3,4-Me <sub>4</sub> C <sub>5</sub>	1.8487	-	-	9.942	9.910
58	C <sub>10</sub>	2.2803	-	-	12.004	12.276
59	C <sub>12</sub>	2.7803	-	-	14.437	14.650
60	C <sub>16</sub>	3.7803	-	-	19.440	19.450

<sup>a</sup>From ref. [24]; <sup>b</sup>With eq. (8); <sup>c</sup>With eq. (9).

tion of TI(3) with  $\Delta H_{\text{vap}}$ . The results are presented in the second part of table 4; again, a set of fifty alkanes was considered for which experimental values of  $\Delta H_{\text{vap}}$  are of highest quality. The regression analysis yielded the quadratic equation (9), a correlation coefficient  $r^2 = 0.9945$  and a standard deviation  $\bar{s} = 0.26 \text{ kcal} \cdot \text{mol}^{-1}$ :

$$\Delta H_{\text{vap}} (\text{kcal} \cdot \text{mol}^{-1}) = 0.0923 \cdot \text{TI}(3)^2 + 4.3977 \cdot \text{TI}(3) + 1.4965. \quad (9)$$

If instead of the parabolic equation (9) a linear regression is attempted, the correlation coefficient decreases to 0.9938, the largest deviation encountered being that for ethane.

Again, this result represents the best correlation of a single TI with  $\Delta H_{\text{vap}}$ , although by using higher-order connectivity indices [24] (multiple TIs) for nearly the same data set (44 alkanes), a correlation coefficient  $r^2 = 0.9999$  and a standard deviation  $\bar{s} = 0.042 \text{ kcal} \cdot \text{mol}^{-1}$  were obtained.

### 5.5. OTHER CORRELATIONS WITH LOIS-BASED TIS

Good correlations were also found between motor octane numbers (MON) of various classes of alkanes (*n*-alkanes, methylalkanes, dimethylalkanes) and TI(3) based on LOISs DSV, AZS and especially ASV. The latter is presented in fig. 7 for *n*-alkanes and all branched alkanes with one methyl or ethyl side chain. Previous correlations of MON with TIs gave better correlation coefficients for a larger set of alkanes [25], not necessarily related structurally.

Other molecular properties were found to correlate satisfactorily with TI(3) based on AZV-LOIS. The properties include liquid phase densities of alkanes (fig. 8), and cavity surface areas of alcohols (fig. 9). These areas in turn were used to correlate nonspecific, local anesthetic properties [26]. The best correlation of partial molal volumes was found for TI(3) based on LOIS ANN for 33 C<sub>5</sub> – C<sub>10</sub> alkanes. For ten *n*-alkanes (C<sub>5</sub> – C<sub>20</sub>), a linear regression yielded for molal volumes  $r^2 = 0.9999$  both *versus* ANN-LOIS and *N*, the number of carbon atoms in the *n*-alkane. Although better correlations with other different TIs have been presented for these properties [24–27], the above examples stress the versatility of the present TIs based on LOISs.

## 6. Conclusions and perspectives

A new approach for graph-theoretical invariants has been presented; it has enabled us to explore twenty new such invariants, but many more can be imagined, and would have to be tested.

On the basis of these new invariants, one may devise a wide variety of topological indices, characterizing the whole graph by a single number. We tested only a few of the possible TIs and found which ones have low degeneracy and good correlational ability. There are as yet a few outstanding examples which surpass all others, but more work is needed before one can reach a definite conclusion and select one optimal TI. A few chemical correlations were tried, with challenging results.

It is highly interesting that the hierarchical ordering of the vertices in graphs provided by a few new graph invariants closely resembles the HOC ordering, which has been shown to correlate with experimental molecular data such as <sup>1</sup>H-NMR chemical shifts in polycyclic aromatic hydrocarbons.

The perspectives which are opened by the new graph invariants are varied both in their mathematical and chemical applications. Thus, one should be able to use extremal values of vertex invariants for defining the graph center or eccentricity. One might devise analogously edge invariants based on the new vertex invariants.

Among possible chemical applications, the first one is to continue and diversify the approach outlined in the present paper. This is being done in our group. Other attractive areas are to determine whether some of the new vertex invariants have any chemical significance; this is no easy task, if one remembers that it needed Erich

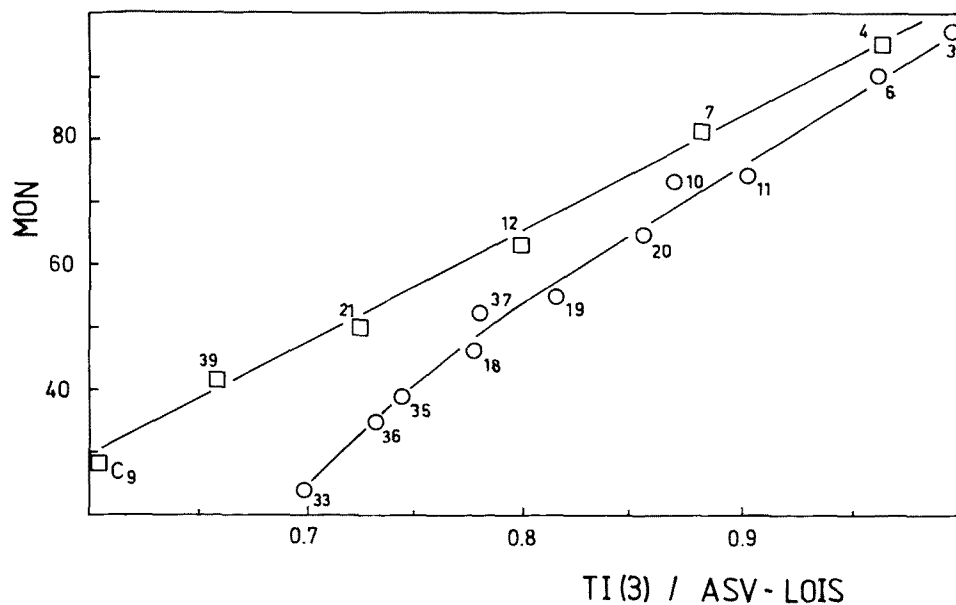


Fig. 7. Dependence of motor octane numbers (MON) of *n*-alkanes ( $\square$ )  $r^2 = 0.9969$  and of branched alkanes with one side chain ( $\circ$ )  $r^2 = 0.9913$ , on  $TI(3)$  computed with ASV-LOIS. Numbers of alkanes are as in table 4.

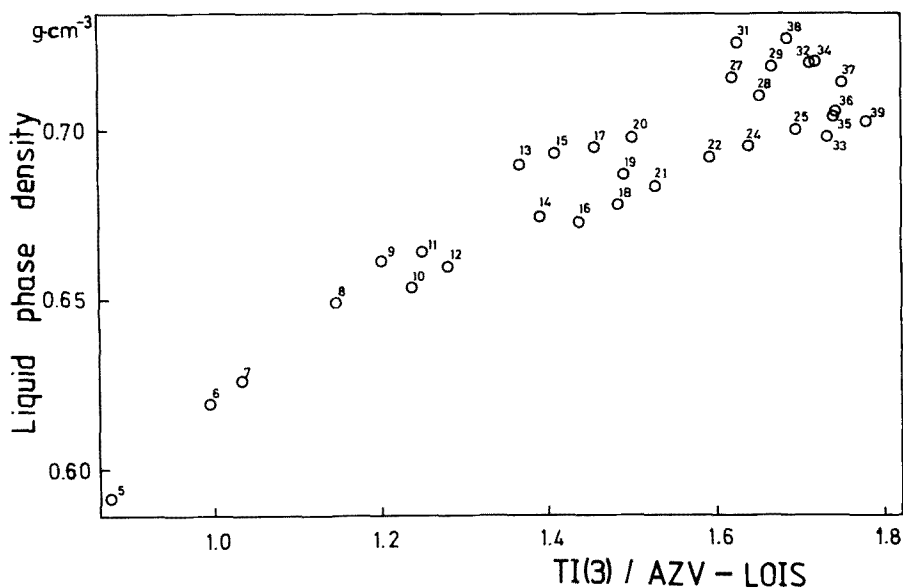


Fig. 8. Dependence of the liquid phase density of alkanes ( $C_5 - C_8$ ) on  $TI(3)$  computed with AZV-LOIS ( $r^2 = 0.904$ ,  $\bar{s} = 0.01 \text{ g} \cdot \text{cm}^{-3}$ ,  $n = 33$ ). Numbers of alkanes are as in table 4.

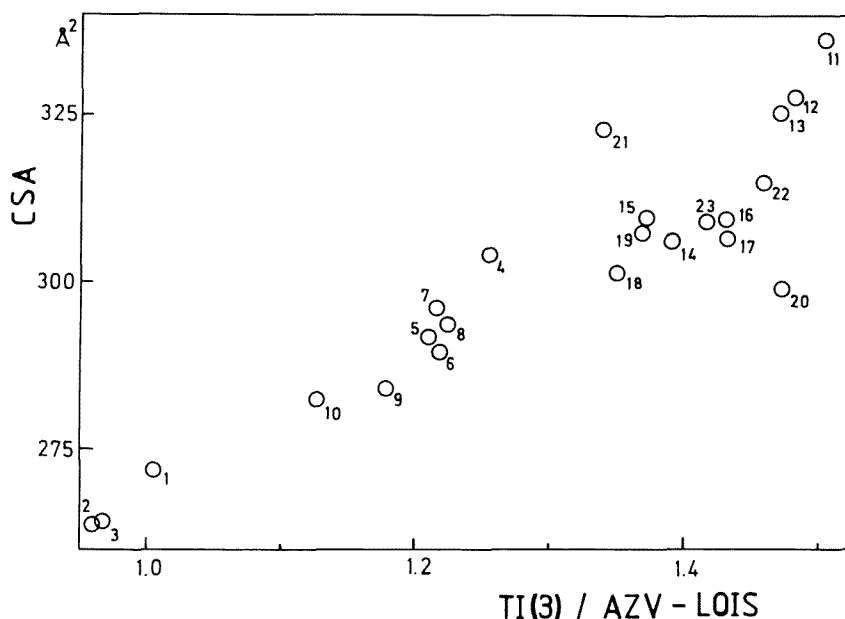


Fig. 9. Dependence of the cavity surface area of alcohols on  $TI(3)$  computed with AZV-LOIS ( $r^2 = 0.8422$ ,  $\bar{s} = 9.6 \text{ \AA}$ ,  $n = 23$ ). Numbers of alcohols are as follows: (1) 1-butanol; (2) 2-methyl-1-propanol; (3) 2-butanol; (4) 1-pentanol; (5) 3-methyl-1-butanol; (6) 2-methyl-1-butanol; (7) 2-pentanol; (8) 3-pentanol; (9) 3-methyl-2-butanol; (10) 2-methyl-2-butanol; (11) 1-hexanol; (12) 2-hexanol; (13) 3-hexanol; (14) 3-methyl-3-pentanol; (15) 2-methyl-2-pentanol; (16) 2-methyl-3-pentanol; (17) 3-methyl-2-pentanol; (18) 2, 3-dimethyl-2-butanol; (19) 3, 3-dimethyl-1-butanol; (20) 3, 3-dimethyl-2-butanol; (21) 4-methyl-1-pentanol; (22) 4-methyl-2-pentanol; (23) 2-ethyl-1-butanol.

Hückel's genius to discover that eigenvalues of adjacency matrices correlated with electronic energies in various orbitals, and that the corresponding eigenvectors (which are vertex graph invariants) represented contributions of atomic orbitals in the LCAO method.

One should also stress the fact that the presence of heteroatoms is taken into account easily and naturally by using atomic numbers  $Z$  as property  $P$  and/or  $R$ ; alternatively, if one wishes to have periodically varying properties, one may use relative electronegativities or relative covalent radii (with carbon as the standard atom), as indicated elsewhere [28]. Previous methods for taking heteroatoms into account [24,29] were fairly unsatisfactory because of their lack of flexibility [29] or of theoretical motivation [24].

## References

- [1] A.T. Balaban, ed., *Chemical Applications of Graph Theory* (Academic Press, London, 1976)
- [2] D.H. Rouvray and A.T. Balaban, in: *Applications of Graph Theory*, ed. R.J. Wilson and L.W. Beineke (Academic Press, New York, 1979).
- [3] N. Trinajstić, *Chemical Graph Theory* (CRC Press, Boca Raton, Florida, 1983) Vols. 1, 2.
- [4] R.B. King, ed., *Chemical Applications of Topology and Graph Theory* (Elsevier, Amsterdam, 1983).
- [5] A.T. Balaban, *J. Chem. Inf. Comput. Sci.* 25(1985)334; *J. Molec. Struct. Theochem.* 120 (1985)117.
- [6] M. Gordon and G.R. Scantlebury, *Trans. Faraday Soc.* 60(1964)605.
- [7] H. Wiener, *J. Amer. Chem. Soc.* 69(1947)17, 2636; *J. Phys. Chem.* 52(1948)425, 1082.
- [8] I. Gutman, B. Rušćić, N. Trinajstić and C.F. Wilcox, *J. Chem Phys.* 62(1975)3399.
- [9] I. Gutman and N. Trinajstić, *Chem Phys. Lett.* 17(1972)535.
- [10] M. Randić, *J. Amer. Chem. Soc.* 97(1975)6609; *Int. J. Quantum Chem. Symp.* 5(1978) 245.
- [11] A.T. Balaban, *Pure Appl. Chem.* 54(1982)1075.
- [12] A.T. Balaban, *Chem. Phys. Lett.* 89(1982)399; A.T. Balaban and P.A. Filip, *Math. Chem.* 16 (1984)163.
- [13] L. Lovasz and J. Pelikan, *Period. Math. Hung.* 3(1973)175.
- [14] H. Hosoya, *Bull. Chem. Soc. Japan* 44(1971)2332; *Int. J. Quantum Chem.* 6(1972)801; H. Hosoya, K. Kawasaki and K. Mizutani, *Bull. Chem. Soc. Japan* 45(1972)3415.
- [15] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press – Wiley, Chichester, 1983); D. Bonchev and N. Trinajstić, *J. Chem. Phys.* 67(1977)4517.
- [16] A.T. Balaban, A. Chiriac, I. Motoc and Z. Simon, *Steric Fit in Quantitative Structure-Activity Relations*, Lecture Notes in Chemistry No. 15 (Springer-Verlag, Berlin, 1980); A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, *Topics Curr. Chem.* 114(1983)21.
- [17] A. Sabljčić and N. Trinajstić, *Acta Pharm. Jugosl.* 31(1981)189.
- [18] A.T. Balaban and L.V. Quintas, *Math. Chem.* 14(1983)213.
- [19] B. Carnahan, H.A. Luther and J.O. Wilkes, *Applied Numerical Methods* (Wiley, New York, 1969) Ch. 5.
- [20] M. Razinger, J.R. Chrétien and J.E. Dubois, *J. Chem. Inf. Comput. Sci.* 25(1985)23.
- [21] O. Mekenyan, D. Bonchev and A.T. Balaban, *J. Comput. Chem.* 5(1984)629; M. Randić, *J. Chem. Phys.* 60(1974)3920; 62(1975)309; *J. Chem. Inf. Comput. Sci.* 17 (1977)171.
- [22] O. Mekenyan, A.T. Balaban and D. Bonchev, *J. Magn. Reson.* 63(1985)1.
- [23] A.T. Balaban, O. Mekenyan and D. Bonchev, *J. Comput. Chem.* 5(1984)629; 6(1985)538.
- [24] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Design* (Academic Press, New York, 1976).
- [25] A.T. Balaban and I. Motoc, *Math. Chem.* 5(1979)197.
- [26] L.B. Kier, L.H. Hall, W.J. Murray and M. Randić, *J. Pharm. Sci.* 64(1975)1971; 65(1976) 1226.
- [27] J.T. Edward, P.G. Farrell and F. Shahidi, *J. Chem. Phys.* 82(1978)2310.
- [28] A.T. Balaban, *Rev. Roum. Chim.*, in press.
- [29] M. Barysz, G. Jashari, R.S. Lall, V.K. Srivastava and N. Trinajstić, in: *Chemical Applications of Topology and Graph Theory*, ed. R.B. King (Elsevier, Amsterdam, 1983) p. 222.