BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN, VOL. 44, 2332—2339 (1971)

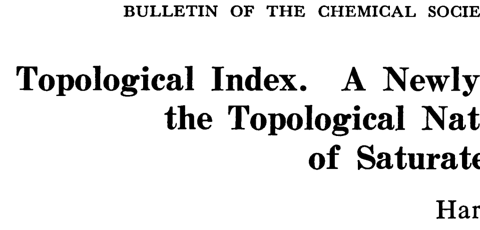
Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons

Haruo Hosoya
Department of Chemistry, Ochanomizu University, Bunkyo-ku, Tokyo

(Received February 27, 1971)

A topological index Z is proposed for a connected graph G representing the carbon skeleton of a saturated hydrocarbon. The integer Z is the sum of a set of the numbers $p(\mathrm{G},k)$, which is the number of ways in which such k bonds are so chosen from G that no two of them are connected. For chain molecules Z is closely related to the characteristic polynomial derived from the topological matrix. It is found that Z is correlated well with the topological nature of the carbon skeleton, *i.e.*, the mode of branching and ring closure. Some interesting relations are found, such as a graphical representation of the Fibonacci numbers and a composition principle for counting Z. Correlation of Z with boiling points of saturated hydrocarbons is pointed out.

There are three conventional methods for distinguishing the carbon skeletons of the structural isomers of saturated hydrocarbons, which may or may not contain a ring or rings as illustrated in Fig. 1.



$$\begin{bmatrix} 0&1&0&0&0&0&0\\ 1&0&1&0&0&1&0\\ 0&1&0&1&0&0&1\\ 0&0&1&0&1&0&0\\ 0&0&0&1&0&0&0\\ 0&1&0&0&0&0&0\\ 0&0&1&0&0&0&0 \end{bmatrix} \qquad \begin{bmatrix} 0&0&1&0&0&0\\ 0&0&1&0&0&0\\ 1&1&0&1&1&0\\ 0&0&1&0&1&0\\ 0&0&1&1&0&1\\ 0&0&0&0&1&0 \end{bmatrix}$$

I II

Fig. 1. Three conventional methods for characterizing the carbon skeletons of saturated hydrocarbons.

i) A systematic naming system in which the size of the ring and the sites of the branching are specified for the longest chain or largest ring.

ii) A graph in which each point corresponds to a carbon atom and the numbers of the attached hydrogen atoms are automatically determined so that each carbon atom satisfies the tetra-valency.

iii) A topological matrix in which the values of the off-diagonal elements 1 and 0 are respectively assinged to the bond and no-bond for the pairs of carbon atoms.

These three methods are equivalent to each other in the topological sense. Topology has contributed to the development of chemistry, *e.g.*, in predicting the numbers of the structural isomers,[1-4] by way of the so-called chemical inference using the structural formulae, through the electronic therories such as the Hückel molecular orbital, valence bond and free electron theories.[5] It has also been shown that some thermodynamic properties such as boiling point are correlated well with the topological nature of the compounds,[6,7] while sophisticated quantum mechanical calculations have been unsuccessful.

In this report a topological index Z is defined and proposed as a fourth candidate for classifying saturated hydrocarbons with respect to their topological nature. Although it does *not* necessarily give the structure of the isomer, it is roughly dependent on the topological

---

[1] E. Cayley, *Ber.*, **8**, 1056 (1875).

[2] H. Schiff, *ibid.*, **8**, 1542 (1875).

[3] H. R. Henze and C. M. Blair, *J. Amer. Chem. Soc.*, **56**, 157 (1934).

[4] G. Polya, *Acta Math.*, **68**, 145 (1938).

[5] Extensive bibliography is given by J. R. Platt, "Encyclopedia of Physics," Vol. 37 (Molecule II), ed. by S. Flügge, Springer-Verlag, Berlin (1961), p. 173.

[6] H. Wiener, *J. Amer. Chem. Soc.*, **69**, 17, 2636 (1947); H. Wiener, *J. Phys. Chem.*, **52**, 425, 1082 (1948).

[7] J. R. Platt, *ibid.*, **56**, 328 (1952).

nature of a molecule, *i.e.*, the size and mode of branching and ring closure. It was also found to correlate well with the boiling point. This magic number Z is enumerated just from the carbon skeleton G of a given saturated hydrocarbon, and is of chemical and mathematical interest.

Theoretical

First, a brief summary of theoretical background is given as premises.[8,9] It seems that most of the formulations and algorithms developed in this paper have not yet been proposed in chemistry or in graph theory.[10]

Premises. A graph G consists of points (vertices or atoms) and lines (edges or bonds). We consider only connected non-directed graphs in which every point has at least one and at most four neighbors, and no loop (a line joining a point to itself) and no multiple lines (double or triple bonds) are involved. With these restrictions the graph is equivalent in a topological sense to the carbon skeleton of a saturated hydrocarbon or, in some cases, an unsaturated hydrocarbon. Distinction between σ- and π-electrons or between single and multiple bonds are meaningless for the moment.

Graphs are grouped into trees and non-trees depending upon whether they contain cycles (rings) or not. Chemically speaking, trees are chain hydrocarbons and non-trees cyclic ones.

An adjacency matrix A for graph G is defined in graph theory as a square matrix with the following elements:

$$a_{ij} = \begin{cases} 1 & \text{if the points } i \text{ and } j \text{ are neighbors,} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The order of A is the number of the points in G. Ham and Ruedenberg[11] called it a topological matrix. Since then a number of interesting relations have been derived.[5] It is obvious that the matrix characters of A are independent of the way of numbering the atoms.

The power series expansion, with respect to X, of the determinant $|XE+A|$ for graph G is called a characteristic polynomial, where E is a unit matrix of the same order N as A.[12]

$$P_G(X) = \det|XE+A| = \sum_{i=0}^{N} f_i X^{N-i} \tag{2}$$

A secular determinant giving the Hückel molecular orbitals for the π-electrons of an unsaturated hydrocarbon is reduced to the same form as above. Besides the standard but iterative methods of Laplace, Newton, and Frame, useful graphical techniques of the expansion of $P_G(X)$ by use of recursion formulae were given by Heilbronner.[13]

Definition of Topological Index. Consider a graph or structure G with N carbon atoms. The topological index Z is defined stepwise as follows.

Definition 1 (*Non-adjacent number*): A non-adjacent number $p(\mathrm{G},k)$ is the number of ways in which k bonds are so chosen from G that no two of them are connected, $p(\mathrm{G},0)$ being unity and $p(\mathrm{G},1)$ the number of the bonds.

Definition 2 (*Z-counting polynomial*):

$$Q_G(Y) = \sum_{k=0}^{m} p(\mathrm{G}, k) Y^k \tag{3}$$

Definition 3 (*Topological index*):

$$Z_G = \sum_{k=0}^{m} p(\mathrm{G}, k) = Q_G(1) \tag{4}$$

where m is the maximum number of k for G.

Note that $p(\mathrm{G},m)$ is equal to the number of Kekulé structures for an unsaturated hydrocarbon in which each atom forming the carbon skeleton G supplies one electron to the π-system.[14] The Z-counting polynomial is introduced as a proof technique and is actually used in Appendix.[16] Unless confusion occurs the suffix G for Z, $P(X)$ and $Q(Y)$ may be omitted.

Let us take two compounds I(tree) and II(non-tree) in Fig. 1 as examples and enumerate these quantities. For convenience the bonds are numbered in italics. For both the compounds $p(\mathrm{G},0)=1$ and $p(\mathrm{G},1)=6$. The following pairs of bonds are the entries of $p(\mathrm{G},2)$:

I: *1-4*, *1-5*, *1-6*, *2-4*, *2-5*, *2-6*, *3-6*, and *4-6*,
II: *1-5*, *1-6*, *2-5*, *2-6*, and *3-6*.

Hence $p(\mathrm{I},2)=8$ and $p(\mathrm{II},2)=5$. For I there are two possible ways of choosing three bonds in which no two of them are connected, *i.e.*, *1-4-6* and *2-4-6*. However, for II one can not choose such combination. We have then $p(\mathrm{I},3)=2$ and $p(\mathrm{II},3)=0$. All other $p(\mathrm{G},k)$ values for $k>3$ are zero and we obtain

$$Q_{\mathrm{I}}(Y) = 1 + 6Y + 8Y^2 + 2Y^3, \qquad Z_{\mathrm{I}} = 17$$
$$Q_{\mathrm{II}}(Y) = 1 + 6Y + 5Y^2, \qquad Z_{\mathrm{II}} = 12.$$

Relation with Characteristic Polynomial. The characteristic polynomials of I and II are obtained as follows by decomposing the secular determinants.

$$P_{\mathrm{I}}(X) = X^7 - 6X^5 + 8X^3 - 2X$$
$$P_{\mathrm{II}}(X) = X^6 - 6X^4 + 2X^3 + 5X^2.$$

Note that for I, i) the power to X decreases by two and ii) the absolute values of the coefficents of $P(X)$ are identical to the set of numbers $p(\mathrm{G},k)$ or the coefficients of the Z-counting polynomial $Q(Y)$. For any tree these two characteristics are observed without exception.[17] i) is proved by the pairing theorem[18]

---

[8] F. Harary, "Graph Theory," Addison-Wesley Publ. Co., Reading, Mass. (1969).

[9] R. G. Busacker and T. C. Saaty, "Finite Graphs and Networks: An Introduction with Applications," McGraw-Hill Book Co., New York, N. Y. (1965).

[10] A complete bibliography of graph theory is given by J. Turner, "Proof Techniques in Graph Theory," ed. by F. Harary, Academic Press, New York, N. Y. (1969), p. 189.

[11] N. S. Ham and K. Ruedenberg, *J. Chem. Phys.*, **29**, 1199 (1958).

[12] Alternatively, one may define $P(X)$ as $\det|XE-A|$.

[13] E. Heilbronner, *Helv. Chim. Acta*, **36**, 170 (1953).

[14] The numbers of Kekulé structures for unsaturated hydrocarbons were discussed by Ham[15] and others. See references cited in Ref. 5.

[15] N. S. Ham, *J. Chem. Phys.*, **29**, 1229 (1958).

[16] In the expansion of $Q_G(Y)$, Y has no significance except for a base symbol to which one can attach exponents.

[17] These relations are approved by term-to-term inspection of the expansion of $\det|XE+A|$ into $P(X)$.

[18] C. A. Coulson and G. S. Rushbrooke, *Proc. Cambridge Phil. Soc.*, **36**, 193 (1940).

for the Hückel molecular orbitals of alternant hydrocarbons, since a tree is alternant. Although ii) appears to be obvious, it seems to have never been stated explicitly. This relation is worthy of being restated as a theorem.

Theorem 1: The characteristic polynomial of a tree T is given by

$$P_{\mathrm{T}}(X) = \sum_{k=0}^{m} (-1)^k p(\mathrm{T}, k) X^{N-2k} \qquad (5)$$

where m is the largest number of bonds disconnected to each other in T.

On the other hand, for non-tree graphs as II the above-mentioned relationship no longer holds, but is contaminated with the terms representing the size and branching mode of the rings.

Normal Paraffins. Denote a normal paraffin with N carbon atoms (n-N-ane) by $\bar{N}$. By induction it can be shown that

$$\begin{aligned} p(\bar{N}, k) &= {}_{N-k}C_k \\ &= (N-k)(N-k-1) \cdots (N-2k+1)/k! \quad (k \neq 0) \qquad (6) \end{aligned}$$

$$P_{\bar{N}}(X) = \sum_{k=0}^{m} (-1)^k {}_{N-k}C_k X^{N-2k} \qquad (7)^{19)}$$

The interger m can be expressed as $m=[N/2]$, where a pair of square brackets are, after Gauss, the greatest integer not exceeding the real number in them.

As an example, for n-heptane one obtains $p(\bar{7},0)=1$, $p(\bar{7},1)=(7-1)/1=6$, $p(\bar{7},2)=(7-2)(7-3)/2\cdot1=10$, $p(\bar{7},3)=(7-3)(7-4)(7-5)/3\cdot2\cdot1=4$, and $p(\bar{7},k)=0$ for $k>3$, and $Z_{\bar{7}}=1+6+10+4=21$, and

$$P_{\bar{7}}(X) = X^7 - 6X^5 + 10X^3 - 4X.$$

From the sets of $p(\bar{N},k)$ the following progression $\{Z_{\bar{N}}\}$ can be obtained.

$$\begin{array}{lllllllll} N: & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \cdots \\ Z_{\bar{N}}: & 1 & 2 & 3 & 5 & 8 & 13 & 21 & 34 \cdots \end{array}$$

The series $\{Z_{\bar{N}}\}$ recurs as

$$Z_{\bar{N}} = Z_{\overline{N-1}} + Z_{\overline{N-2}} \qquad (N \geq 3) \qquad (8)$$

This is just the Fibonacci series[20] and is given by the Binét formula[21]

$$Z_{\bar{N}} = \left\{ \left(\frac{1+\sqrt{5}}{2} \right)^{N+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{N+1} \right\} \Big/ \sqrt{5} \qquad (9)$$

Relation (8) can be proved from the definition of the topological index and the well-known recursion formula for $P_{\bar{N}}(X)$:[13]

$$P_{\bar{N}}(X) = X \cdot P_{\overline{N-1}}(X) - P_{\overline{N-2}}(X) \qquad (N \geq 3) \qquad (10)$$

Branched Paraffins. For branched paraffins $P(X)$ obtained from Eq. (5) without tedious effort of decomposing the determinant $|XE+A|$. The results for the lower members of normal and isomeric paraffins are given in Table 1 in the increasing order of the topological

TABLE 1. $p(\mathrm{G},k)$ VALUES AND TOPOLOGICAL INDEX Z FOR THE GRAPHS CORRESPONDING TO ALL THE POSSIBLE ISOMERS OF THE LOWER MEMBERS OF SATURATED HYDROCARBONS

a. $N=1\sim6$

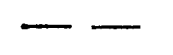
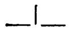
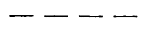
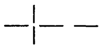
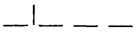
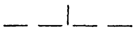
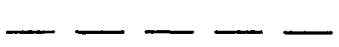
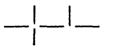
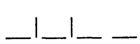
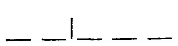
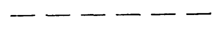
N	Graph (G)	$p(\mathrm{G},k)$[a] $k=0$	1	2	3	Z_{G}	bp(°C)[b]
1		1	0	0	0	1	-161.7
2		1	1	0	0	2	-88.6
3		1	2	0	0	3	-42.2
4		1	3	0	0	4	-11.7
		1	3	1	0	5	-0.5
5		1	4	0	0	5	9.5
		1	4	2	0	7	27.9
		1	4	3	0	8	36.1
6		1	5	3	0	9	49.7
		1	5	4	0	10	57.9
		1	5	5	0	11	60.2
		1	5	5	1	12	63.5
		1	5	6	1	13	68.7

TABLE 1. b. $N=7$

Graph (G)	$p(\mathrm{G},k)$[a] $k=0$	1	2	3	Z_{G}	bp(°C)[b]
	1	6	6	0	13	80.9
	1	6	7	0	14	79.2
	1	6	8	0	15	80.5
	1	6	7	2	16	86.0
	1	6	8	2	17	89.7
	1	6	9	2	18	90.0
	1	6	9	3	19	91.9
	1	6	9	4	20	93.4
	1	6	10	4	21	98.4

---

[19] In other words we have

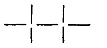
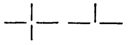
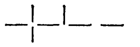
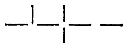
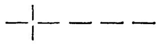
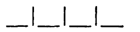
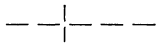
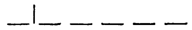
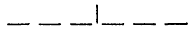
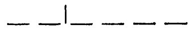
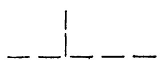
$$Q_{\bar{N}}(Y) = \sum_{k=0}^{m} {}_{N-k}C_k Y^k. \qquad (7')$$

[20] The author thanks Prof. E. Heilbronner, Basel, for his pointing out the Fibonacci series. The series $\{Z_{\bar{N}}\}$ rapidly converges to $\{(1+\sqrt{5})/2\}^{N+1}/\sqrt{5}$. For computer calculation one may use the expression with Gaussian brackets as

$$Z_{\bar{N}} = \left[\left(\frac{1+\sqrt{5}}{2} \right)^{N+1} \Big/ \sqrt{5} + 0.5 \right]. \qquad (9')$$

[21] N. N. Vorob'ev, "Fibonacci Numbers," Nauka Publ., Moscow (1964) (Japanese translation by T. Tsutsui, Tokyo Tosho Publ. Co., Tokyo (1966)).

Table 1. c. $N=8$

Graph (G)	$p(G,k)^{a)}$ $k=0$	1	2	3	4	Z_G	bp(°C)$^{b)}$
	1	7	9	0	0	17	106.3
	1	7	11	0	0	19	99.3
	1	7	11	3	0	22	109.9
	1	7	11	4	0	23	114.8
	1	7	12	3	0	23	106.9
	1	7	12	4	0	24	113.5
	1	7	13	4	0	25	109.1
	1	7	12	5	0	25	112.0
	1	7	13	5	0	26	109.5
	1	7	13	6	0	27	115.6
	1	7	13	7	0	28	115.7
	1	7	12	7	1	28	118.3
	1	7	13	7	1	29	117.7
	1	7	14	7	0	29	117.7
	1	7	14	8	0	30	117.7
	1	7	14	8	1	31	119.0
	1	7	14	9	1	32	118.6
	1	7	15	10	1	34	125.7

a) The characteristic polynomial $P(X)$ and Z-counting polynomial $Q(Y)$ are respectively given by Eqs. (5) and (3).

b) American Petroleum Institute Research Project 44 at the National Bureau of Standards, "Selected Values of Physical and Thermodynamic Properties of Hydrocarbons Tables No. 1a, 2a, 3a, and 4a (1945).

index Z. Note the numbering character of Z. All the possible isomeric saturated hydrocarbons from methane to n-hexane are found to be "numbered" with Z except for neopentane which has the same Z value 5 as n-butane leaving the compound-6 blank. The Z value of 2,2,3-trimethylbutane, 13, is the same as that of n-hexane and the nine isomers of heptane are nicely "numbered" up to 21. For eighteen isomers of octane, numbering from 17 to 34 is a little disordered and redundancy increases very smoothly with N, since it is an outcome of a different way of partitioning a given interger Z into a set of $p(G,k)$'s.

Let us take a closer look at the $p(G,k)$ values. To all the isomers with the same number of carbon atoms the values of $p(G,0)$ and $p(G,1)$ are common by difinition. If one defines the decrement of the $p(G,k)$ value of an isomeric N-ane relative to the normal N-ane as

$$\delta p(N,k) = p(\bar{N},k) - p(N,k), \tag{11}$$

we find additive contributions to the value of $\delta p(N,2)$ by the branching as

$$\delta p(N,2) = \text{(Number of tertiary carbon atoms)} + 3\text{(Number of quarternary carbon atoms)}. \tag{12}$$

Although higher members of $\delta p(N,k)$'s have similar additive relations, fairly complicated crowdness corrections are necessary, *e.g.*, terms which are functions of the number of adjacent pairs of tertiary and quarternary carbon atoms. Anyway a general trend of the lowering of the resultant Z value by the branching is thus evident.

Composition Principle for Z. In order to study the above problem quantitatively, the composition principle for enumerating the Z values of branched hydrocarbons is given as a theorem.

Theorem 2: The Z value of a given graph G is calculated as

Z_G = (Product of the Z values for the graphs G_1 and G_2 derived from G by cutting a bond b)
+ (Product of the Z values for all the graphs G_{i1}, G_{i2}, ⋯ ($i=1$, 2) derived from G_1 and G_2 by cutting all the bonds incident to b in the *original* graph G).[22]

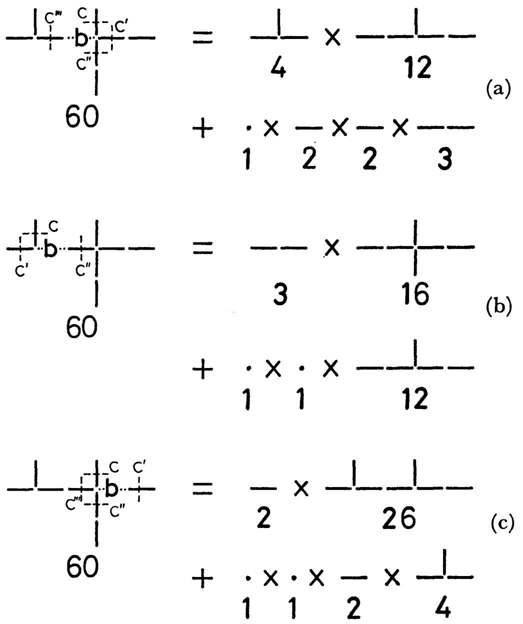


Fig. 2. Examples of the composition principle for the topological index Z. Bond b is cut first and then bonds c,c',.... Three different choices of b give the same Z value.

---

[22] For ring compounds one can choose a bond whose deletion gives rise to one graph G_1 (=G-b) instead of generating two fragmental graphs. In this case read the first term in Theorem 2 as (The Z value of the graph G_1 derived from G by cutting a bond b).

See Fig. 2 for explanation. Graph G may be called 2,4-dimethyl-4-ethylhexane. Choose any one bond say bond b in Fig. 2a and cut it. The graph G is divided into isobutane and 3-methylpentane, whose Z values are, respectively, 4 and 12, giving the product 48. Then cut all the bonds c, c′, c″, and c‴ which *were* incident to bond b in the original graph G. The resultant graphs are one methane ($Z=1$), two ethanes ($Z=2\times 2$) and one propane ($Z=3$), giving the product of Z's as 12. The Z value of the graph G is the sum of 48 and 12, namely 60. It should be confirmed that the Z value is uniquely obtained independent of the choice of bond b in the first step. This is actually the case as illustrated in Figs. 2b and 2c.

If one cuts the terminal bond of n-N-ane and applies Theorem 2, the recursion formula (8) is obtained, while if the jth bond is chosen the following equation is obtained.

$$Z_{\overline{N}} = Z_{\overline{j}}\cdot Z_{\overline{N-j}} + Z_{\overline{j-1}}\cdot Z_{\overline{N-j-1}} \qquad (13)$$ [23]

A general proof of Theorem 2 is given in the Appendix. It is also shown that this composition principle can be applied to ring compounds (*vide infra*). If we neglect the second term of the right-hand side of Theorem 2, and if necessary apply Theorem 2 successively, we get the following theorem.

Theorem 3: If graph G is decomposed into a set of fragmental subgraphs G_1, G_2, $\cdots$by deleting a set of bonds b, c,$\cdots$($G-b-c-\cdots=G_1+G_2+\cdots$, $G_1\cap G_2=0$, $\cdots$), then

$$Z_G > Z_{G_1}Z_{G_2}\cdots.$$

Especially for a n-N-ane we get

$$Z_{\overline{N}} > Z_{\overline{i}}\cdot Z_{\overline{N-i}}$$
$$Z_{\overline{N}} > Z_{\overline{i}}\cdot Z_{\overline{j}}\cdot Z_{\overline{N-i-j}} \qquad (14)$$
$$Z_{\overline{N}} > Z_{\overline{i}}\cdot Z_{\overline{j}}\cdot Z_{\overline{k}}\cdot Z_{\overline{N-i-j-k}} \quad \text{and the like.}$$

The ralations are derived also from Eq. (13). By the aid of these composition principles let us study the regularities in the lowering of Z values by branching.

First consider a series of 2-methyl-($N-1$)-ane (Fig. 3a, $j=2$), in which a branch of a unit length is attached to the second atom counting from one end of a linear chain composed of ($N-1$) atoms. Application of Theorem 2 to this graph gives the general expression

$$Z_{2m-N} = Z_{\overline{N-1}} + Z_{\overline{N-3}} \quad (N\geq 3) \qquad (15)$$

where $2m$-N stands for the 2-methyl substituted isomer of N-ane. Generally for j-methyl substituted isomers we get

$$Z_{jm-N} = Z_{\overline{N-1}} + Z_{\overline{j-1}}\cdot Z_{\overline{N-j-1}} \quad (N-1\geq j\geq 2) \qquad (16)$$ [24]

Similarly for j,j-dimethyl-($N-2$)-ane ($N-2\geq j\geq 2$,

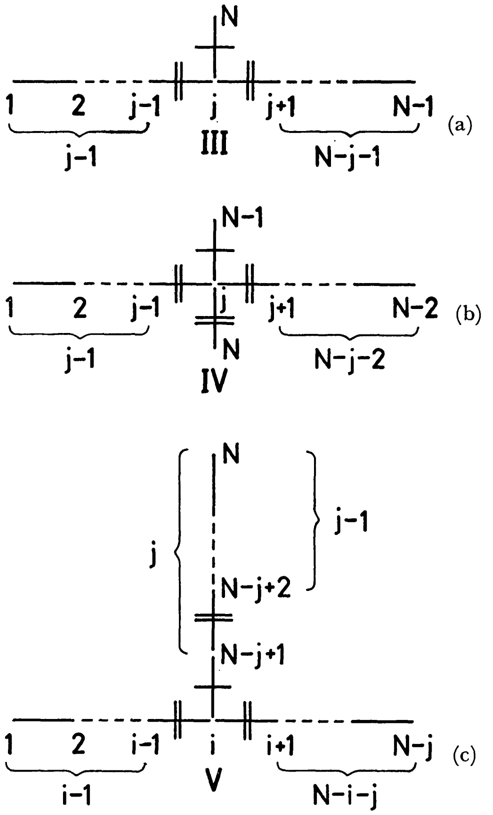
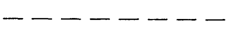
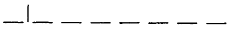
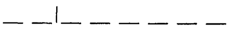
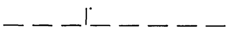
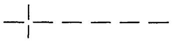
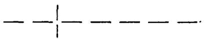
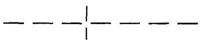
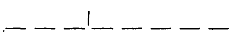
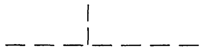
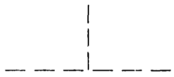


Fig. 3. Application of the composition principle to three different classes of branched hydrocarbons. A single line and two parallel lines dissecting bonds, represent the first and second cutting processes, respectively.

TABLE 2. EFFECT OF BRANCHING ON THE $p(G,k)$ AND Z VALUES OF DECANE ISOMERS

Type[a]	j	Graph (G)	$p(G,k)$[b] $k=2$	3	4	5	Z_G
n-decane			28	35	15	1	89
III	2		27	30	9	0	76
	3		27	31	12	1	81
	4		27	31	11	0	79
	5		27	31	11	1	80
IV	2		25	22	3	0	60
	3		25	24	7	0	66
	4		25	24	5	0	64
V	1		27	31	11	0	79
	2		27	32	13	1	83
	3		27	32	12	0	81

a) See the corresponding structures in Fig. 3 with N=10 and $i=4$ (for V).

b) For all the graphs $p(G,0)=1$ and $p(G,1)=9$.

---

[23] This is also proved by the known relation for $P(X)$,[13] *viz.*,

$$P_{\overline{N}}(X) = P_{\overline{j}}(X)P_{\overline{N-j}}(X) - P_{\overline{j-1}}(X)P_{\overline{N-j-1}}(X) \qquad (13')$$

or $$Q_{\overline{N}}(Y) = Q_{\overline{j}}(Y)Q_{\overline{N-j}}(Y) + Y\cdot Q_{\overline{j-1}}(Y)Q_{\overline{N-j-1}}(Y) \qquad (13'')$$

[24] The corresponding expressions for the Z-counting and characteristic polynomials are as follows.

$$Q_{jm-N}(Y) = Q_{\overline{N-1}}(Y)+Y\cdot Q_{\overline{j-1}}(Y)Q_{\overline{N-j-1}}(Y) \qquad (16')$$
$$P_{jm-N}(X) = X\cdot P_{\overline{N-1}}(X)-P_{\overline{j-1}}(X)P_{\overline{N-j-1}}(X) \qquad (16'')$$
$$Q_{jjm-N}(Y) = Q_{\overline{N-2}}(Y)+2Y\cdot Q_{\overline{j-1}}(Y)Q_{\overline{N-j-2}}(Y) \qquad (17')$$
$$P_{jjm-N}(X) = X\cdot P_{\overline{N-2}}(X)-2\cdot P_{\overline{j-1}}(X)P_{\overline{N-j-2}}(X) \qquad (17'')$$
$$Q_{ji-N}(Y) = Q_{\overline{i}}(Y)Q_{\overline{N-i}}(Y)+Y\cdot Q_{\overline{i-1}}(Y)Q_{\overline{j-1}}(Y)Q_{\overline{N-i-j}}(Y) \qquad (18')$$
$$P_{ji-N}(X) = X\cdot P_{\overline{i}}(X)P_{\overline{N-i}}(X)-P_{\overline{i-1}}(X)P_{\overline{j-1}}(X)P_{\overline{N-i-j}}(X) \qquad (18'')$$

denoted by *jjm-N*) series, we get (Fig. 3b)

$$Z_{jjm-N} = Z_{\overline{N-2}} + 2Z_{\overline{j-1}} \cdot Z_{\overline{N-j-2}} \quad (N-2 \geq j \geq 2) \qquad (17)^{24)}$$

The result for the compound in which a chain of j atoms is attached to the ith position of n-$(N-j)$-ane is as follows.

$$Z_{ij-N} = Z_{\bar{j}} \cdot Z_{\overline{N-j}} + Z_{\overline{i-1}} \cdot Z_{\overline{j-1}} \cdot Z_{\overline{N-i-j}} \quad (N-j \geq i \geq 1) \qquad (18)^{24)}$$

Examples of the calculated results are shown in Table 2 for decane isomers ($N=10$). Note the zig-zag variation of the Z values within a series of compounds and an interesting pattern of the $p(\mathrm{G},k)$ values. By successive applications of Eqs. (8), (13), and (14) to Eqs. (16)—(18), the following inequality is obtained.

$$Z_{\bar{N}} > Z_{jm-N},\ Z_{jjm-N},\ Z_{ji-N} \qquad (19)$$

More generally this can be stated as a theorem.

Theorem 4: The value Z of the normal compound is the largest among the isomers with the same number of carbon atoms.

Table 3. Topological properties of cyclo-N-anes

N	$P(X)$	$Q(Y)$	Z
3	X^3-3X+2	$1+3Y$	4
4	X^4-4X^2	$1+4Y+2Y^2$	7
5	X^5-5X^3+5X+2	$1+5Y+5Y^2$	11
6	$X^6-6X^4+9X^2-4$	$1+6Y+9Y^2+2Y^3$	18
7	$X^7-7X^5+14X^3-7X+2$	$1+7Y+14Y^2+7Y^3$	29
8	$X^8-8X^6+20X^4-16X^2$	$1+8Y+20Y^2+16Y^3+2Y^4$	47

Cycloparaffins. For compounds with one or more rings, simple correlation between the $p(\mathrm{G},k)$'s and the coefficients of the characteristic polynomials is not always observed. $P(X)$, $Q(Y)$, and Z for the lower members of cyclo-N-ane (denoted by N^0) are given in Table 3. The coefficeints in $Q(Y)$ are systematically obtainable as

$$\begin{aligned} Q_{N^0} &= \sum_{k=0}^{m} ({}_{N-k}C_k \cdot N/(N-k))Y^k \\ &= 1 + NY + N(N-3)/2 \cdot Y^2 \\ &\quad + N(N-4)(N-5)/3! \cdot Y^3 \\ &\quad + N(N-5)(N-6)(N-7)/4! \cdot Y^4 + \cdots \qquad (20) \end{aligned}$$

For ring compounds, however, no one-to-one correspondence between the coefficients of $P(X)$ and $Q(Y)$ exists.

For Z_{N^0} values the following two relations are found.

$$Z_{N^0} = Z_{\bar{N}} + Z_{\overline{N-2}} \qquad (N \geq 3) \qquad (21)$$

$$Z_{N^0} = Z_{N-1^0} + Z_{N-2^0} \qquad (N \geq 5) \qquad (22)$$

Note that Eq. (21) is a special case of Theorem 2 and the recursion formula (22) can be derived from Eq. (21). The series $\{Z_{N^0}\}$ form another family of the Fibonacci series, which can be expressed as

$$Z_{N^0} = \left(\frac{1+\sqrt{5}}{2}\right)^N + \left(\frac{1-\sqrt{5}}{2}\right)^N \qquad (23)^{25)}$$

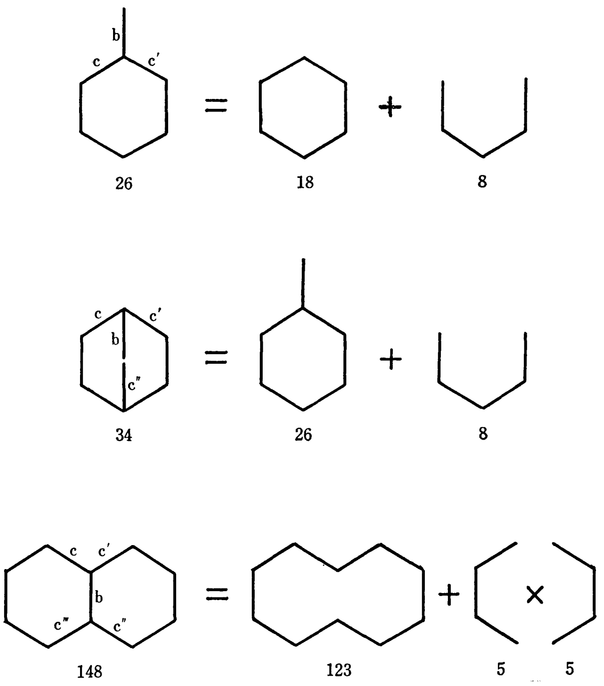


Fig. 4. Examples of application of the composition principle to ring compounds. The letters b, c,···have the same meaning as in Fig. 2.

Table 4. Topological properties of the graphs corresponding to all the possible isomers of the lower members of monocyclic saturated hydrocarbons

N	Number[a)]	Z	n_i[b)]				bp(°C)[c)]
			$i=4$	3	2	1	
3	1	4	0	0	3	0	−32.8
4	2	6	0	1	2	1	0.7
	3	7	0	0	4	0	12.5
5	4	8	1	0	2	2	20.6
	5	9	0	2	1	2	32.6
	6	10	0	1	3	1	35.9
	7	10	0	1	3	1	36.3
	8	11	0	0	5	0	49.3
6	9	12	1	1	1	3	52.4
	10	13	1	0	3	2	56
	11	14	1	0	3	2	56.8
	12	14	0	3	0	3	62.9
	13	14	0	2	2	2	58.3
	14	14	0	2	2	2	59
	15	15	0	2	2	2	62.8
	16	15	0	2	2	2	64
	17	16	0	1	4	1	69.2
	18	16	0	1	4	1	71.8
	19	17	0	1	4	1	70.6
	20	18	0	0	6	0	80.7

a) The numbers refer to the graphs in Chart 1.
b) The number of the carbon atoms with i neighboring carbon atoms. Namely, the numbers of quarternary, tertiary, secondary, and primary carbon atoms.
c) The data were taken from the reference in Table 1.

The composition principle (Theorem 2) can also be applied to complicated cycloparaffin derivatives as illustrated in Fig. 4. In Table 4 are given the topological indices of all the possible isomers of mono-

---

25) For computer calculation one can use the Gaussian expression.

$$Z_{N^0} = \left[\left(\frac{1+\sqrt{5}}{2}\right)^N + 0.5\right]. \qquad (23')$$

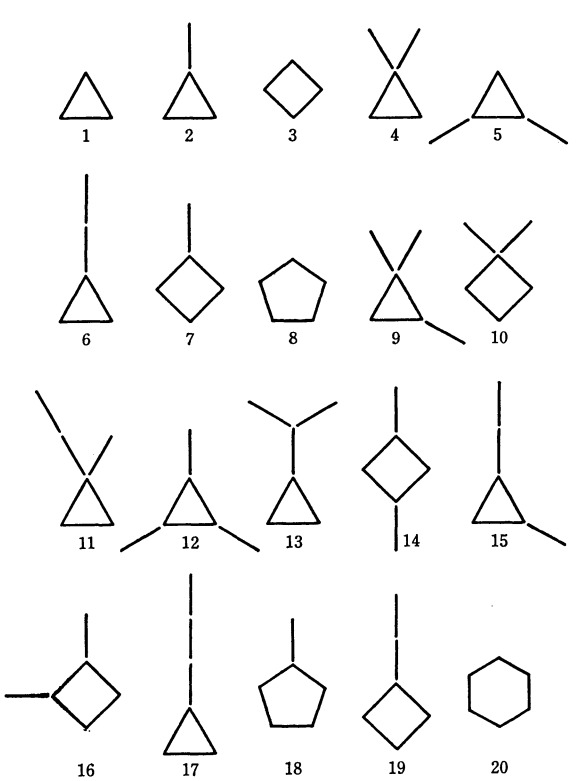


Chart 1

cyclic hydrocarbons with the carbon numbers from three to six.

The numbers of quarternary, tertiary, secondary, and primary carbon atoms for each graph are also shown. It is interesting to note that if one reads a set of the four integers for each graph as a four-digit integer it decreases as the Z value increases as long as the isomers of the same carbon number are compared. This suggests, though qualitatively, that the topological index for ring compounds also decreases with branching. This sort of calculation can be extended to higher members of the series and polycyclo compounds.[26]

Discussion

Comparison with Other Studies. Wiener proposed a path number w (Wiener number), which is the total number of bonds between all pairs of carbon atoms.[6] For example the w value of compound I in Fig. 1 is 46. This quantity turns out to be the half sum of the off-diagonal elements of a distance matrix $\mathbf{D}$ whose element d_{ij} is a number of bonds for the shortest path between atoms i and j.[27] Although it increases with carbon number and decreases with branching, no systematic relations such as recursion formula and composition principle, are found except for normal paraffins ($w=(n^3-n)/6$).

In ralation to the Wiener number, Altenburg proposed a polynomial $\sum_i n_i a_i$ for a carbon skeleton where n_i is the number of pairs of atoms whose distance is a_i or i C–C bonds.[28] The expression for compound I is $6a_1+7a_2+6a_3+2a_4$.[29]
Note that

$$w = \sum_i i \cdot n_i \tag{24}$$

For I we get $w=6\times1+7\times2+6\times3+2\times4=46$. As in the case of w, it is difficult to find useful relations for this polynomial expression.

Relation to Boiling Points. It was found that the boiling points of a number of saturated hydrocarbons are correlated fairly well with Z values. It is well known that for *non*-polar molecules the heavier the molecule the higher the boiling point becomes and the more the molecule is branched the lower the boiling point. The Z value has these characteristics. A number of empirical equations have been proposed to estimate the boiling points of a series of structural isomers by use of structural parameters and other physical quantities.[6,7,28,30,31] However, all of them require from three to six parameters and tedious calculations. In our case only one parameter is required. We can show that many empirical rules on the relation between boiling points of saturated hydrocarbons and structures of carbon skeleton are proved if we admit an empirical equation relating the Z value and boiling point. Of the physical and thermochemical quantities boiling point seems to have the best correlation with the topological index. It is also worth mentioning that of the quantities studied boiling point is the only quantity whose upper or lower limit for the isomeric compounds is the value of the normal compound.

It must be remembered that the topological index is defined for a topological graph which has nothing to do with distances, angles, steric strain, or hindrance. Therefore, without substantial modification, the value Z alone can not account for the difference in physical quantities of geometrical isomers such as *cis*- and *trans*-1,2-dimethylcyclopropane, nor for the overcrowded effect expected for compounds having vicinal quarternary carbon atoms. In this sense the proposed quantity Z is a kind of "index" for problems which are expected to be dependent on the topological nature of the skeleton of a system like a molecule.

Appendix

Proof of Theorem 2. Given three bonds a, b, and c and three graphs A, B, and C, a larger graph L is constructed as in Fig. 5a. It may be expressed as

$$\mathrm{L} = (\mathrm{A}+\mathrm{a}) + (\mathrm{B}+\mathrm{b}) + (\mathrm{C}+\mathrm{c}).$$

Similarly let us have

$$\mathrm{M} = (\mathrm{D}+\mathrm{d}) + (\mathrm{E}+\mathrm{e}) + (\mathrm{F}+\mathrm{f}).$$

---

[26] K. Mizutani, K. Kawasaki, and H. Hosoya, *Natural Science Report of Ochanomizu Univ.*, Tokyo, **22**, 39 (1971).

[27] Since Wiener did not treat ring compounds this is an extention of his definition.

[28] K. Altenburg, *Kolloid-Z.*, **178**, 112 (1961).

[29] Instead of a_i, a symbol a_i^2 is used in Ref. 28 meaning the root-mean-square value of the distance between all the atom pairs separated by i bonds. However, in the present paper we only consider topological distances.

[30] K. Li, R. L. Arnett, M. B. Epstein, R. B. Ries, L. P. Bitler, J. M. Lynch, and F. D. Rossini, *J. Phys. Chem.*, **60**, 1400 (1956); J. B. Greenshields and F. D. Rossini, *ibid.*, **62**, 271 (1958).

[31] K. Altenburg, *Brennstoff-Chem.*, **47**, 100, 331 (1966).

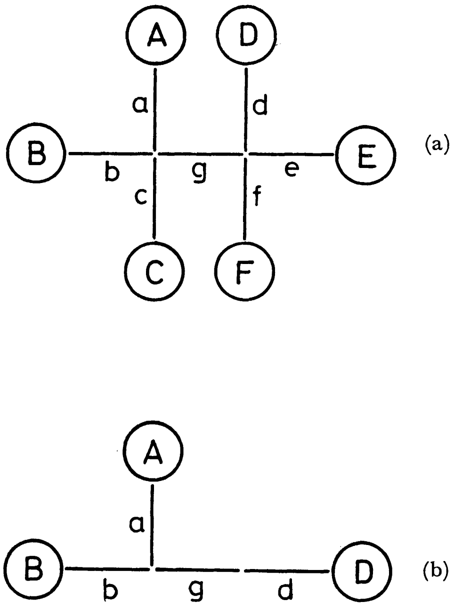


Fig. 5. Explanation for the proof of Theorem 2. The subgraphs A, B,···F are disconnected with each other.

By joining the graphs L and M with a bond g, a graph G can be constructed:

$$\mathrm{G} = \mathrm{L} + \mathrm{g} + \mathrm{M}$$

Now consider a number $p(\mathrm{G},k)$ for the graph G, the number of ways in which such k bonds are chosen from G that no two of them are connected. Each combination either includes or excludes the bond g. Since Z_G is the sum of $p(\mathrm{G},k)$'s, it is also divided into two parts, namely, g-including and g-excluding. It may be formulated as

$$Z_\mathrm{G} = Z_\mathrm{G}(+\mathrm{g}) + Z_\mathrm{G}(-\mathrm{g}). \tag{A1}$$

Let us expand the following polynomial product,

$$\begin{aligned} Q_\mathrm{L}(Y)Q_\mathrm{M}(Y) &= \sum_{j=0}^{m_\mathrm{L}}\sum_{k=0}^{m_\mathrm{M}} p(\mathrm{L},j)p(\mathrm{M},k)Y^{j+k} \\ &= (0,0)+\{(0,1)+(1,0)\}Y \\ &\quad + \{(0,2)+(1,1)+(2,0)\}Y^2 + \cdots \\ &\quad + \{\cdots+(m_\mathrm{L}, m-m_\mathrm{L})+\cdots \\ &\quad + (m-m_\mathrm{M}, m_\mathrm{M})+\cdots\}Y^m \end{aligned} \tag{A2}$$

where the following abbreviated symbol is adopted:

$$(j, k) = p(\mathrm{L},j)p(\mathrm{M},k). \tag{A3}$$

The cofficient of Y^i in thr far right-hand side of Eq. (A2) is equal to the number of ways in which such i bonds are chosen from the graph G−g (or L+M) that no two of them are conneceted. Hence

$$Z_\mathrm{G}(-g) = Q_\mathrm{L}(1)Q_\mathrm{M}(1) = Z_\mathrm{L}Z_\mathrm{M}. \tag{A4}$$

On the other hand, the coefficients in the following product of the polynomials *multiplied* by Y are the contributions of the g-including combinations.

$$\begin{aligned} Q_\mathrm{ABC}(Y)Q_\mathrm{DEF}(Y)Y &= \sum_{j=0}^{m_\mathrm{ABC}}\sum_{k=0}^{m_\mathrm{DEF}} p(\mathrm{ABC},j)p(\mathrm{DEF},k)Y^{j+k+1} \\ &= (0,0)Y + \{(0,1)+(1,0)\}Y^2 \\ &\quad + \{(0,2)+(1,1)+(2,0)\}Y^3 \\ &\quad + \cdots + \{\cdots+(m_\mathrm{ABC}, m-m_\mathrm{ABC}-1)+\cdots\}Y^m, \end{aligned} \tag{A5}$$

where

$$\mathrm{ABC} = \mathrm{L} - (\mathrm{a}+\mathrm{b}+\mathrm{c}),$$

$$\mathrm{DEF} = \mathrm{M} - (\mathrm{d}+\mathrm{e}+\mathrm{f}),$$

and

$$(j, k) = p(\mathrm{ABC},j)p(\mathrm{DEF},k). \tag{A6}$$

Thus we obtain

$$\begin{aligned} Z_\mathrm{G}(+\mathrm{g}) &= \{Q_\mathrm{ABC}(Y)Q_\mathrm{DEF}(Y)Y\}_{Y=1} \\ &= Q_\mathrm{ABC}(1)Q_\mathrm{DEF}(1). \end{aligned} \tag{A7}$$

Since the subgraphs A, B, and C in ABC are not connected to each other, the polynomial $Q_\mathrm{ABC}(Y)$ is equal to the product of the constituent $Q(Y)$'s.

$$Q_\mathrm{ABC}(Y) = Q_\mathrm{A}(Y)Q_\mathrm{B}(Y)Q_\mathrm{C}(Y) \tag{A8}$$

Similarly we have

$$Q_\mathrm{DEF}(Y) = Q_\mathrm{D}(Y)Q_\mathrm{E}(Y)Q_\mathrm{F}(Y). \tag{A9}$$

Therefore we obtain

$$\begin{aligned} Z_\mathrm{G}(+\mathrm{g}) &= Q_\mathrm{A}(1)Q_\mathrm{B}(1)Q_\mathrm{C}(1)Q_\mathrm{D}(1)Q_\mathrm{E}(1)Q_\mathrm{F}(1) \\ &= Z_\mathrm{A}Z_\mathrm{B}Z_\mathrm{C}Z_\mathrm{D}Z_\mathrm{E}Z_\mathrm{F} \end{aligned} \tag{A10}$$

and

$$Z_\mathrm{G} = Z_\mathrm{L}Z_\mathrm{M} + Z_\mathrm{A}Z_\mathrm{B}Z_\mathrm{C}Z_\mathrm{D}Z_\mathrm{E}Z_\mathrm{F}. \tag{A11}$$

Note that in the derivation no special restriction was imposed on the number of the subgraphs incident to the bond g. Thus for the system G′ as in Fig. 5b we obtain

$$Z_{\mathrm{G}'} = Z_{\mathrm{L}'}Z_{\mathrm{M}'} + Z_\mathrm{A}Z_\mathrm{B}Z_\mathrm{D} \tag{A12}$$

where

$$\mathrm{L}' = (\mathrm{A}+\mathrm{a}) + (\mathrm{B}+\mathrm{b})$$

$$\mathrm{M}' = \mathrm{D} + \mathrm{d}$$

$$\mathrm{G}' = \mathrm{L}' + \mathrm{g} + \mathrm{M}'.$$

In this case, however, Eq. (A12) is already implied by Eq. (A11) if we consider graphs C, E, and F to be null and assume their Z values to be unity. This is consistent with the relation $Z_{\bar{0}}=1$ obtained from Eq. (9).

Extension of this theorem into ring compounds is approved if we note that a ring is formed by connecting any two of the subgraphs A, B,···F in the graph G.[22)]