Topological Centric Coding and Nomenclature of Polycyclic Hydrocarbons. I. Condensed Benzenoid Systems (Polyhexes, Fusenes)

2 AUTHORS:

Danail Bonchev
Virginia Commonwealth University
273 PUBLICATIONS **4,282** CITATIONS

Alexandru T. Balaban
Texas A&M University - Galveston
688 PUBLICATIONS **9,375** CITATIONS

When there are two centers in a given tree, it can be coped with by inserting a dummy vertex between the two vertexes, converting the case of one center (see Figure 9).

CONCLUSION

A method to generate chemical structures is presented. It is based on BCT representation of chemical structures, labeling nodes of bipartite graphs (BCT) corresponding to blocks, and allocating cutpoints in the blocks.

The programs were implemented in FORTRAN on a ACOS-900 computer, and the spectral data were edited from EPA/NIH Mass Spectral Database (1975 edition, about 11 300 records). BCT file is organized according to parameters m,n as shown in Table I. The execution speed of structure generation depends on the times of orbit computation in the generation stages such as BCT labeling, cutpoint assignment, and connecting blocks. Apparently orbits should be computed for each node of a generation tree except leaves, so the number of orbits computation equals the number of nodes (not leaves) of generation trees. The execution time of orbit computation depends on the number of vertexes and the symmetric degree of the graph, so we prepared three types of programs for orbit

computation. This method of structure generation is useful for automatic structure elucidation and constitutes a part of ASASMAS.

REFERENCES AND NOTES

(1) Nelson, D. B.; Munk, M. E.; Gash, K. B.; Herald, D. L., Jr. *J. Org. Chem.* **1969**, *34*, 3800.
(2) Antosz, F. J.; Nelson, D. B.; Herald, D. L., Jr.; Munk, M. E. *J. Am. Chem. Soc.* **1970**, *92*, 4933.
(3) Shelly, C. A.; Munk, M. E. *Anal. Chim. Acta* **1978**, *103*, 245.
(4) Shelly, C. A.; Munk, M. E. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 247.
(5) Kudo, Y.; Sasaki, S. *J. Chem. Doc.* **1974**, *14*, 200.
(6) Kudo, Y.; Sasaki, S. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 43.
(7) Masinter, L. M.; Sridharan, N. S.; Lederberg, J.; Smith, D. H. *J. Am. Chem. Soc.* **1974**, *96*, 7702.
(8) Masinter, L. M.; Sridharan, N. S.; Carhart, R. E.; Smith, D. H. *J. Am. Chem. Soc.* **1974**, *96*, 7714.
(9) Carhart, R. E.; Smith, D. H.; Brown, H.; Djerassi, C. *J. Am. Chem. Soc.* **1975**, *97*, 5755.
(10) Carhart, R. E.; Smith, D. H.; Brown, H.; Sridharan, N. S. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 124.
(11) Nakayama, T.; Fujiwara, Y. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 23.
(12) Harary, F. "Graph Theory"; Addison–Wesley: Reading, MA, 1969.
(13) Busacker, R. G.; Saaty, T. L. "Finite Graphs and Networks: An Introduction with Applications"; McGraw-Hill: New York, 1965; Chapter 6.
(14) Aho, A. V.; Hopcroft, J. E.; Ullman, J. D. "The Design and Analysis of Computer Algorithm"; Addison–Wesley: Reading, MA, 1974.

Topological Centric Coding and Nomenclature of Polycyclic Hydrocarbons. 1. Condensed Benzenoid Systems (Polyhexes, Fusenes)

DANAIL BONCHEV*

Higher School of Chemical Technology, Department of Physical Chemistry, Burgas, Bulgaria

ALEXANDRU T. BALABAN

Organic Chemistry Department, The Polytechnic, Bucharest, Roumania

On the basis of the focal numbering of each benzenoid ring in the graphite lattice, any polyhex (cata-, peri-, and corona-fused polycyclic benzenoid system) can be coded by making use of (i) the dualist graph of the polyhex and (ii) the recent generalization of the topological center concept. A vertex of the dualist graph [which vertex is the (or one of the) generalized (oligo) center vertex (vertexes)] functions as the focal point, and the coding of the polyhex follows unambiguously from a few simple conventions. The numerical code can be used for nomenclature purposes, and the present approach can be easily generalized to condensed polycyclic nonbenzenoid systems as will be shown in future publications.

INTRODUCTION

The coding and nomenclature of condensed benzenoid polycyclic hydrocarbons (polyhexes),[1,2] planar conjugated condensed nonbenzenoid systems,[3] bridged saturated polycyclic systems,[4] and diamondoid hydrocarbons[5] were examined in previous publications. The unsatisfactory present rules for the nomenclature of polyhexes were analyzed and discussed earlier.[1,2] Since the present IUPAC rules (and even the newly proposed nodal nomenclature[6]) do not have a topological foundation but are based on the longest chain of alkanes, or longest string of condensed benzenoid rings, the resulting nomenclature is very cumbersome. A topological solution based on dualist graphs (see below) was proposed,[1,2] but it was limited to cata-condensed benzenoid systems. So far no satisfactory treatment of peri- or corona-condensed benzenoid systems has been found.

On the basis of the ideas of dualist graphs and generalized topological centers of these dualist graphs, we propose a novel

approach which encompasses all planar totally conjugated systems; the present paper deals with benzenoid systems. Future publications will examine bidimensional (planar) nonbenzenoid hydrocarbons and will generalize the present approach to tridimensional diamondoid hydrocarbons.

It has long been known that a tree (an acyclic graph) has a topological invariant which is a vertex (center) or a pair of adjacent vertexes (bicenter). The use of the center or bicenter of an acyclic graph for a topological nonconventional nomenclature system for alkanes was advocated by Read[7] and Lederberg et al.[8] However, until now, this approach was limited to noncyclic graphs.

## DUALIST (CHARACTERISTIC) GRAPHS

A benzenoid condensed polycyclic hydrocarbon (a polyhex or a fusene) having n condensed benzenoid rings contains sp^2-hybridized carbon atoms and (with the notable exception of helicenes with $n \geq 6$) is a portion of the bidimensional

224 *J. Chem. Inf. Comput. Sci., Vol. 21, No. 4, 1981*

BONCHEV

graphite lattice. If the centers of the benzenoid rings are marked by points (vertexes) and if two such points are joined by a line (edge) whenever the corresponding benzenoid rings are condensed (fused) by having two common carbon atoms, then we obtain the *dualist* (or characteristic) graph of the polyhex. Such graphs are similar to the dual graphs encountered in graph theory, but unlike the latter their angles are fixed, and there is no vertex corresponding to the outer region. The dualist graphs were used[1,2] for a new definition of cata vs. peri condensation: dualist graphs of cata-condensed benzenoid systems (catafusenes) are trees (have no cycles), while those of peri- or corona-condensed systems have cycles. If the cycles are all three-membered, the system is peri condensed (perifusene), and if larger rings appear in the dualist graph the system is corona condensed (coronafusene). If both three-membered and larger rings appear in the dualist graph, the condensation is mixed. [Classically, corona condensation is considered to exist when *only* larger rings than three-membered ones appear in the dualist graph. Since, however, mixed peri + cata condensation is assimilated to peri condensation, we consider that it is convenient to ignore the italicized *only* in the preceding phrase, and to assimilate mixed corona + peri, corona + cata, or corona + peri + cata condensation to corona condensation. Thus there is a hierarchical order of precedence: corona, peri, cata condensation.] In the new definition, all catafusenes with the same number of rings *n* are isomeric, and they include helicenes irrespective of the *n* value. The newly prepared kekulene is an example of a coronafusene.[9] By means of the dualist graph, enumerations of nonbranched and branched[10] catafusenes were carried out, and proposals for their coding and nomenclature were made.[1,2] By denoting linear annulation with 0 and angular annulation with 1 or 2, a three-digit code was developed which specifies not only the structure of any catafusene but can be also correlated[11] with resonance energies (through the number of 0 digits) and carcinogenicities (through the number of bay regions).[12]

TOPOLOGICAL CENTERS OF GRAPHS

The concept of the center of a graph[13] is essentially based on the notions of paths and distances in the graph. Different paths can exist that start from one vertex and end in another one along edges of the graph so that all vertexes (and thus necessarily all edges) are distinct. The shortest path between two vertexes (measured topologically in terms of the number of edges it contains) is said to be their (topological) distance. The distances in the graph depicting a certain hydrogen-depleted molecule can be determined simply by counting the number of bonds (edges) connecting each pair of atoms (vertexes) along the shortest path. A concise expression of all distances of a given vertex is the so-called *vertex* (or *atomic*) *distance code*.[14] The distances d_i from the given vertex to all other vertexes in the graph are written in this code as an increasing sequence of numbers having as superscripts the number of times they appear in the graph: $d_1{}^a d_2{}^b d_3{}^c \ldots d_{max}{}^k$. All topological paths of length l_i starting from each vertex of the given graph are similarly expressed by the so-called *vertex* (*atomic*) *path code*:[15] $l_1{}^d l_2{}^e l_3{}^f \ldots l_{max}{}^j$. As an example, the paths of length 2 for vertex 1 in graph I include the following vertexes: 132, 134, 142, 143, and 152, i.e., five paths of length two denoted as 2^5 in the path code. The paths of length three of vertex 1 are 1324, 1325, 1342, 1432, 1423, 1425, 1523, and 1524, i.e., 3^8 in the path code. The paths of length four are 13425, 14325, 15234, and 15243, i.e., 4^4 in the path code. Of course, vertex 2 is equivalent to vertex 1; therefore, it will have the same path code. No paths of length higher than four may exist in graph I, since any edge added to a path of length four leads to a cycle(s), contradicting thereby the definition of a path. For acyclic graphs, the distance code coincides with the

Table I. Example of Distance and Path Codes, Distance Radius, and Distance Rank in a Graph
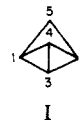
| vertex | distance | | | path code |
	code	radius	rank	
1	$1^3 2^1$	2	5	$1^3 2^5 3^8 4^4$
2	$1^3 2^1$	2	5	$1^3 2^5 3^8 4^4$
3	$1^3 2^1$	2	5	$1^3 2^6 3^6 4^6$
4	$1^3 2^1$	2	5	$1^3 2^6 3^6 4^6$
5	$1^3 2^1$	2	6	$1^3 2^4 3^8 4^8$

path code, but for cyclic graphs this is no longer true. The vertex distance and path codes of graph I are indicated in Table I.
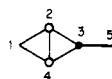
According to the classical definition,[16-18] the center of a graph is constituted by the vertex or vertexes whose maximum distance d_{max} to any other vertex is minimal; in this case, d_{max} is called the radius r_d of the graph. As seen from Table I, in graph I all five vertexes should thereby be classified as center since their radii are the same ($d_{max} = r_d = 2$). This example illustrates the low discrimination ability of the classical definition for graph center in the case of cyclic or polycyclic molecules.

For an optimal coding and nomenclature of cyclic molecules, the center of the corresponding graph should be more precisely defined so as to reduce the number of vertexes in the center as much as possible. Such a generalized definition of graph center was recently proposed by us.[13] The procedure for determining the graph center includes the requirement for a minimal d_{max} (the classical definition) but only as a first criterion. When several nonequivalent vertexes have the same radius, the second criterion selects those vertexes which have the smallest distance rank d_r (the sum of all distances of this vertex). For graph I, vertexes 1–4 have the same distance rank $d_r = 5$, while for vertex 5 the value $d_r = 6$ is larger; hence this vertex 5 is eliminated from the search of the graph center. The third criterion in the procedure deals with a situation in which several vertexes have the same radius and distance rank but are at variance in their distance codes. For instance, let us assume the following distance codes for two vertexes in graphs whose other vertexes are not specified:

graph A   graph B
vertex 1: $1^2 \ 2^4 \ 3^1$ vertex 7: $1^2 \ 2^3 \ 3^2 \ 4^1$
vertex 2: $1^3 \ 2^2 \ 3^2$ vertex 4: $1^3 \ 2^1 \ 3^3 \ 4^1$

In such cases, those vertexes are selected as centers in whose distance code the largest distance appears a minimum number of times (vertex 1 for graph A). If the largest distance occurs the same number of times for several vertexes, as in graph B, one should compare stepwise the next lower distance, etc. Thus for graph B, vertex 7 is selected. The center obtained by applying criteria 1–3 can consist of one or several vertexes; we call it the *pseudocenter* of the graph.[13]

A further discrimination between the pseudocenter vertexes is provided by the fourth criterion. The subgraph (consisting only of the pseudocenter vertexes and the edges joining them as in the initial graph) is consecutively treated by criteria 1–3, and if necessary, the process is repeated until on two successive cycles no further vertex deletion is obtained. The resulting "center" is either one single vertex or composed of several vertexes which may or may not be connected. This resulting "center" is named the *polycenter* of the graph.

CODING AND NOMENCLATURE OF POLYCYCLIC HYDROCARBONS

J. Chem. Inf. Comput. Sci., Vol. 21, No. 4, 1981 **225**

Table II. Path Code and the Finding of the Oligocenter of a Graph

II

vertex	path code
2	$1^3 2^5 3^3 4^1$
4	$1^3 2^5 3^3 4^1$
3	$1^3 2^4 3^4$

Returning to the example of graph I presented in Table I, its pseudocenter consists of a graph with four vertexes which are not discriminated by criteria 1–3. The distance codes of these four vertexes are


vertexes 1 and 2: $1^2 \ 2^1$
vertexes 3 and 4: 1^3

Therefore, according to criterion 4, we apply again criterion 1 which decides that vertexes 3 and 4 constitute the polycenter of graph I. The radius of vertexes 3 and 4 in the pseudocenter is 1, which is smaller than the radius 2 of vertexes 1 and 2. Thus the search for the center of graph I ends in two equivalent vertexes, 3 and 4. As seen in this example, on applying criteria 1–4, one reduces substantially the number of vertexes in the graph center by comparison with the classical definition. At each step, before applying the next criterion, one should check the topological equivalence of the vertexes remaining in the graph center (two atoms in a molecule are regarded as topologically equivalent if their interchange does not alter the adjacency relationships). If all vertexes in the graph center prove to be equivalent, the search is terminated, even if some criteria had not yet been employed, i.e., the equivalence check helps to discontinue the search at an earlier stage in favorable cases.

On the other hand, in more complicated graphs, even after using criteria 1–4, not all of the vertexes in the polycenter may be equivalent. In such cases, one may apply a refined procedure for graph center determination[19] which we developed recently. It consists of four analogous criteria, denoted by 1′–4′, based on the vertex path code instead of the vertex distance code. These four additional criteria are consecutively applied to the subgraph formed by the polycenter vertexes, and as a result, they may decrease even further the number of vertexes in the center. The final result from the application of criteria 1–4 and 1′–4′ consists of, in practically all cases, topologically equivalent vertexes, which constitute the *oligocenter* of the graph. This oligocenter is actually the generalized center of the graph.

For instance, graph II, presented in Table II, is found after application of criteria 1–4 to have a polycenter of three vertexes (denoted by O and ●). It must be stressed that the path codes refer to the initial graph, not to the polycenter vertexes only. Indeed, in the initial graph the polycenter vertexes are topologically nonequivalent, but in the polycenter subgraph they appear equivalent and could not be discriminated by path codes which only take into account the vertexes of the polycenter. As seen from the path codes presented in Table II, no path of length 4 occurs for vertex 3, whose *path radius* is minimal, $l_r = 3$. Therefore, criterion 1′ specifies vertex 3 as the oligocenter consisting of one vertex. One more illustration will be presented for criterion 2′ (Table III). The polycenter of graph III consists of two vertexes. Their path radii prove to be the same, but they differ in their *path ranks*, so the oli-

Table III. Path Code and Path Rank for Finding the Oligocenter of a Graph
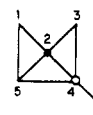
III

vertex	path code	path radius	path rank
2	$1^4 2^7 3^6 4^3$	4	$1 \times 4 + 2 \times 7 + 3 \times 6 + 4 \times 3 = 48$
4	$1^4 2^6 3^7 4^3$	4	$1 \times 4 + 2 \times 6 + 3 \times 7 + 4 \times 3 = 49$
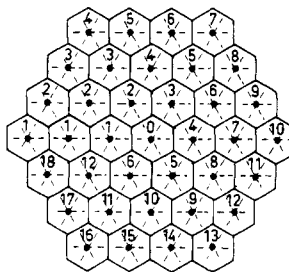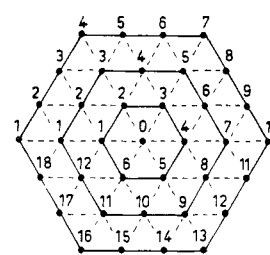
(a)   (b)

Figure 1. (a) Standard numbering of the benzenoid rings in the graphite (honeycomb) lattice around the focal point. (b) Successive shells around the focal point and the standard numbering of vertexes of dualist graphs.

gocenter is vertex 2 (in Tables I–III the polycenter vertexes are marked by O while the oligocenter vertexes are marked by ●; O vertexes are discarded on going from the polycenter to the oligocenter).

Usually, however, criteria 1–4 are sufficient for determining the center (oligocenter), and thus the more complicated path codes (criteria 1′–4′) are seldom used.

CENTRIC CODING OF POLYHEXES (FUSENES)

Since we are able to find the center point(s) of the dualist graph in a polyhex, we may consider such a central point as a *focal point* (denoted by 0) of a shell structure. We apply the procedure described in the previous section to locate the center of the dualist graph by using criteria 1–4 or, if necessary, also criteria 1′–4′. If the oligocenter consists of more than one vertex, any of the vertexes belonging to the oligocenter may be selected as the focal point; the choice will be made unique by the condition that the code be minimal. One exception should be mentioned here, though it will be discussed in more detail later. For helicenes, corona-condensed systems, or systems with mixed (cata and/or peri + corona) condensation, the focal point need not be a vertex belonging to the oligocenter but may be an inner (virtual) vertex, which does not correspond to an existing benzenoid ring.

We propose a standard focal numbering of each ring in the graphite (honeycomb) lattice, as indicated in Figure 1a,b. In Figure 1a, the possible C–C bonds of the polyhex are indicated by full lines; in Figure 1, a and b, the possible edges of the dualist graph are indicated by broken lines. The numbering in each shell around the focal point begins from the benzenoid ring whose center lies on the left of the focal point on the horizontal line and continues clockwise. For the *i*th shell, the rings or, correspondingly, the vertexes of the dualist graph are thus numbered from 1–6*i*. One has to imagine the same numbering system as in Table I continuing in all directions to infinity, by successive addition of more and more concentric shells around the focal point. [We have also considered, but rejected, another focal numbering system with collinear points of the dualist graph having the same numbers and with re-
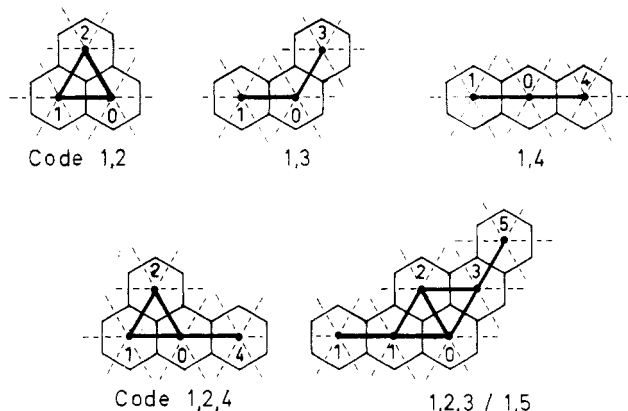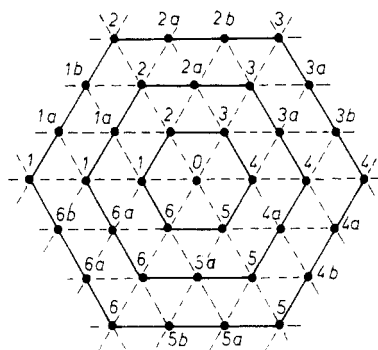
Figure 2. Two catafusenes and three perifusenes superimposed on the standard honeycomb lattice in the orientation giving the focal code.

maining points in each shell being denoted by additional letters as seen in the accompanying drawing.

Thus the following examples of codes would have been used in place of the system described in the main text (the corresponding correct codes are given in brackets): 1,3/1a,2a,3a (= 1,3/2,4,6), 1,3/1a,3a/3a (= 1,3/2,6/8), and for a corona-condensed system, /1-3,5/3a,4,4a,5a,6,6a (= /1-3,5/6-8,10-12). Despite some obvious advantages (a simple means of finding rectilinear portions of the duualist graph), the greater drawbacks (less concise notation, needs additional rules for alphanumerical ordering) pleaded against this system.]

To obtain the code of any polyhex (fusene), the following procedure is proposed:

(1) The polyhex is depicted so that some C–C bonds of the regular hexagons (benzenoid rings) appear vertically.
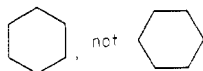
The dualist graph of the polyhex is obtained (some of its edges will therefore be horizontal, and none of its edges will be vertical), and the center of the dualist graph is found as described in the preceding section.

(2) The unique vertex of the (oligo) center, or one of the (oligo) center vertexes, is denoted by a zero, and a line of the dualist graph emerging from this focal point 0 is placed in horizontal position toward the left. The second end point of this horizontal line is denoted by 1. This ensures a superposition of the given polyhex (fusene) on the standard honeycomb lattice from Figure 1 in such a manner as to have always 1 as the first digit in the code. The other vertexes in the dualist graph are numbered according to Figure 1, i.e., according to the number of the vertex of the standard honeycomb lattice on which they become superimposed.

(3) The coding starts from the first shell surrounding the focal point. For the brevity of the code, the focal point is not mentioned explicitly, with the exception of benzene [another exception will be mentioned in item 6] which is coded as 0,

because it only has one vertex in the dualist graph. The numberings of vertexes in the first shell are listed in increasing order and separated by commas. Thus, naphthalene is coded as 1, and the three trihexes are coded as 1,2 (perinaphthenyl radical), 1,3 (phenanthrene), and 1,4 (anthracene), as seen in Figure 2. If a longer string of consecutive digits appears, it may be contracted to the first and last one separated by a hyphen; instead of 1,2,3,4,5 one may write 1-5. If there exist vertexes of the dualist graph in more remote shells, a slash (/) is added, and the numberings in the next shell are listed in increasing order, and so on.

(4) For the code to be unique, the sequence of digits, ignoring commas and slashes, must correspond to the smallest number among all possible choices of the lines emerging from the focal point or of the focal point among the vertexes of the oligocenter. If numbers higher than 9 appear, one has to consider a numbering system in a higher base than 10 so that each point of the dualist graph could be considered as denoted by one digit. Additional precedence rules making use of slashes will be given in the next (5th) item to define an ordering of fusenes.

In order to find the minimal code, all possible orientations of the fusene around the focus (namely, for asymmetrical polyhexes three orientations differing by 120°, followed by another three orientations after turning the polyhex on the other face) and all assignments of the focus to the oligocenter vertexes (namely, as many assignments as classes of topologically nonequivalent vertexes in the oligocenter) should be tested. With a little practice, the most favorable orientation and assignment are easily found. The last two polyhexes in Figure 2 serve as examples; according to rules 1–4 above, the code for benzoperinaphthenyl is 1,2,4; the other five possible orientations result in higher numbers, namely, 1,2,5; 1,3,4; 1,3,6; 1,4,5; and 1,4,6. For the last example in Figure 2, where again only the orientation has to be established since the center is one vertex, dibenzo[d,l]pyrene has code 1,2,3/1,5 rather than 1,5,6/1,9.

Evidently, in finding the center of the dualist graph, no attention is paid to its angles. However, once the (oligo) center has been found, angles of the dualist graph must be preserved.

(5) For ordering polyhexes with the same number of benzenoid rings in a table, a precedence order is established according to the hierarchy of shells which are delineated by slashes. Proceeding as in item 4 above, but ignoring only commas and comparing the content of each shell as outlined by slashes, an inner shell has precedence over an outer shell, while within the same shell precedence is ruled first by the larger number of vertexes and secondly by the smaller values of the numberings of an equal number of vertexes. Thus for hexahexes ($n = 6$) a priority order is, for example, 1,2,3,4,5 > 1,2,3,4/1 > 1,2,4,5/2 > 1,2,4/1,3 > 1,3/1,4,6 > 1,3/1,4/6. These precedence rules are used in Tables IV and V. Table IV presents the fusenes with $n = $ 1–5, together with their dualist graphs and their focal codes. Table V presents the dualist graphs of the hexahexes (hexafusenes), i.e., the case $n = 6$, together with their focal codes. Since we consider the smallest helicene to be hexahelicene and since helicenes are coded as will be indicated in the following item, Table V contains hexahelicene in brackets.

(6) The above five rules are sufficient for coding cata- and perifusenes. For coronafusenes, helicenes with six or more rings, or systems with mixed condensation (peri and/or cata + corona), as an exception to the above rules, the lattice vertexes of the standard lattice within the larger rings of the dualist graph, or the lattice vertex surrounded by the helicene rings, can also function as focal points, though these *virtual vertexes* do not correspond to existing (actual) benzenoid rings and hence do not belong to the oligocenter. In these cases,

Table IV. Perifusenes and Catafusenes with $n = 1$–5, Their Dualist Graphs, Focal Points, and Focal Codes

n	Polyhexes
1	0
2	1
3	1,2 1,3 1,4
4	1,2,3 1,2,4 1,3,5 1,3 / 1 1,3 / 2 1,3 / 6 1,4 / 1
5	1,2,3,4 1,2,3,5 1,2,4,5 1,2,5 / 2 1,2,4 / 1 1,2,4 / 3 1,2,4 / 4 1,2,4 / 6 1,2,4 / 7 1,2,4 / 8 1,3,5 / 1 1,3,5 / 2 1,3 / 1,4 1,3 / 1,5 1,3 / 1,6 1,3 / 2,4 1,3 / 2,6 1,3 / 6,12 1,4 / 1,6 1,4 / 1,7 1,4 / 2,6 1,4 / 2,8

Table V. Dualist Graphs, Focal Points, and Focal Codes of Fusenes with $n = 6$

1,2,3,4,5	1,2,4,5 / 3	1,2,4 / 4,7
1,2,3,4 / 1	1,2,4,5 / 4	1,2,4 / 4,8
1,2,3,4 / 2	1,2,4 / 1,3	1,2,4 / 6,7
1,2,3,4 / 3	1,2,4 / 1,6	1,2,4 / 6,8
1,2,3,4 / 4	1,2,4 / 1,7	1,2,4 / 6,12
1,2,3,4 / 8	1,2,4 / 1,8	1,2,4 / 7,8
1,2,3,5 / 3	1,2,4 / 2,6	1,2,4 / 7,12
1,2,3,5 / 5	1,2,4 / 2,7	1,2,4 / 8,12
1,2,3,5 / 6	1,2,4 / 2,8	1,3,5 / 1,4
1,2,3,5 / 8	1,2,4 / 3,6	1,3,5 / 1,5
1,2,3,5 / 9	1,2,4 / 3,7	1,3,5 / 1,6
1,2,3 / 1,5	1,2,4 / 3,8	1,3,5 / 2,4
1,2,3 / 5,12	1,2,4 / 3,12	1,3,5 / 2,6
1,2,4,5 / 2	1,2,4 / 4,6	1,3,5 / 2,8
1,3,5 / 2,12	1,3 / 1,4 / 1	1,3 / 2,6 / 2
1,3 / 1,2,4	1,3 / 1,4 / 2	1,3 / 2,6 / 8
1,3 / 1,2,5	1,3 / 1,4 / 5	1,3 / 6,12 / 8
1,3 / 1,2,6	1,3 / 1,4 / 6	1,4 / 1,2,6
1,3 / 1,3,4	1,3 / 1,4 / 18	1,4 / 1,2,7
1,3 / 1,4,6	1,3 / 1,5 / 1	1,4 / 1,2,8
1,3 / 1,4,12	1,3 / 1,5 / 2	1,4 / 1,6,8
1,3 / 1,5,6	1,3 / 1,5 / 8	1,4 / 2,6,8
1,3 / 1,6,7	1,3 / 1,6 / 1	1,4 / 1,6 / 1
1,3 / 1,6,12	1,3 / 1,6 / 2	1,4 / 1,6 / 2
1,3 / 2,3,4	1,3 / 1,6 / 8	1,4 / 1,6 / 18
1,3 / 2,4,6	1,3 / 1,6 / 9	1,4 / 1,7 / 1
1,3 / 2,6,7	1,3 / 1,6 / 18	
1,3 / 2,6,12	1,3 / 2,4 / 2	

two types of condensation into corona condensation.

the choice of the focal point is determined by the same rules 1–4, considering, however, as the dualist graph a "virtual dualist graph", including not only the actual vertexes of the dualist graph but also all inner virtual vertexes. Thus a compact notation is obtained following a nonsubjective determination of the focal point and the minimal code. Whenever the determination of the focal point is done in this way, i.e., by considering the virtual dualist graph, the code starts with a slash. A code starting with slash(es) indicates also that the focal point is a *virtual focus* unless the digit zero appears in the code; namely, in mixed peri- and corona-condensed systems like the last example in Table VI where the focus results on an actual vertex of the virtual dualist graph, this focus has to be indicated by the digit 0. Helicenes with $n = 6$, 7, 10, and 13 are coded by /1-6, /1-6,1, /1-6,1-4, and /$(1\text{-}6)^2$,1, respectively.

If the numbering starts in the first shell around a virtual focus, the code starts with one slash. When the first shell around the virtual focus is empty, the fact that the code refers to the second shell is signaled by a double slash at the beginning of the code, as seen in Table VI. For larger coronaphenes having also their second, third, etc., shells around the virtual focus empty, a corresponding larger number of slashes is used at the beginning of the code.

Table VI presents a few systems with corona condensation, with mixed condensation, including hexahelicene, and with virtual or actual focal points, together with their codes. It is seen that the present treatment considers together corona and mixed condensation so that it is convenient to assimilate these

DISCUSSION

The proposed coding system is applicable to all benzenoid polycyclic systems: catafusenes, perifusenes, and coronafusenes. A special group of catafusenes are the helicenes, whose dualist graphs contain (starting with heptahelicene) repeating numbering of rings (vertexes of the dualist graph) within the same shell. Though helicenes, being catafusenes, might be

Table VI. Five Coronafusenes, One Fusene with Mixed (Peri + Corona) Condensation, and Hexahelicene; Their Dualist Graphs and Focal Codes are Presented. Larger Point Indicates the Virtual (or Actual, for the Last Graph) Focal Point

/ 1 - 5 / 10-12   / 1,2,4,5 / 4 - 6,10-12   / 1,2,3,5 / 6 - 8,10 -12   / 1 - 8

hexahelicene

/ 1 - 4 / 8-12   // 1 -12   / 0 / 1 - 4 / 8 -12
kekulene
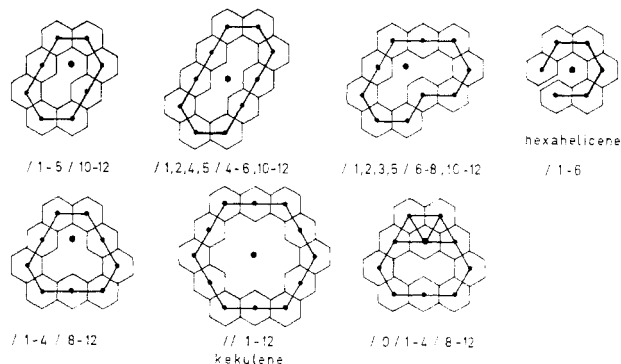
coded like these, with actual foci, we prefer their coding with virtual foci.

The number n of condensed benzenoid rings can easily be retrieved from the code; if the code does not start with slash(es), i.e., if the system is cata (with the exception of helicenes) or peri condensed, n is one plus the number of symbols (numberings) in the code. If the code starts with a slash, n equals the number of symbols (numberings) in the code, and the system is either a helicene, or has corona or mixed (corona + peri) condensation with a virtual focal point, unless an actual focal point is signaled by a zero following the initial slash when the preceding rule applies.

Our search revealed no instance where a fusene could not afford a minimal code by unique selection of one vertex from the oligocenter vertexes. Though we do not exclude the possibility that with complicated fusenes two different oligo-center vertexes could lead to identical codes without being topologically equivalent, we consider such a case very unlikely.

As indicated earlier,[1] unlike catafusenes or coronafusenes which are normal aromatics (unless they are too crowded like the lower coronaphenes with strong nonbonded interactions between the inner hydrogens), perifusenes or systems with mixed condensation can be normal like perylene (code 1,2,4,5) or pyrene (1,2,3), monoradicals like perinaphthenyl (1,2), diradicals (1,2,3,4/4), triradicals (1-6/2,6,10), compounds with localized double bonds like zethrene (1,2,4/7,8), etc.

Given a code, the construction of the dualist graph (hence of the fusene) is extremely simple; the vertexes are marked on Figure 1a or 1b, and whenever two vertexes appear in vicinal positions they are joined by a line to form the dualist graph. The opposite process, namely, to find the code of a given polyhex, requires some practice. Therefore we plan to prepare more extensive tables for fusenes to supplement those presented here in Tables IV–VI.

Recently, a detailed study of molecular cyclicity in polycyclic condensed hydrocarbons was carried out[20] on the basis of the topological distances in the graph. The sum of these distances, called *distance number* or *Wiener number*[21-25] proved to be a reliable reverse proportional measure of molecular cyclicity. On this basis the essence of molecular cyclicity was expressed in a number of rules reflecting the influence of different structural factors. Many molecular properties obey these rules, such as π-electron energy, delocalization energy, heat of formation, boiling point, half-wave polarographic reduction potential, etc. This makes possible a relative ordering of molecules according to a certain rule, related to a given property. Quantitative correlations with the Wiener number are also easily obtained. Proceeding from the idea that the focal code proposed in the present paper is also based on distances in the graph, one might expect it to have features similar to the Wiener number. The examples given in Figure

Code :   1,2   1,3   1,4

$\bar{a}$: $\dfrac{3}{9}$ $\dfrac{2}{9}$ $\dfrac{2}{9}$

Code :   1,2,3   1,2,4   1,3 / 2   1,4 / 1

$\bar{a}$: $\dfrac{10}{24} = \dfrac{5}{12}$ $\dfrac{8}{24} = \dfrac{4}{12}$ $\dfrac{6}{24} = \dfrac{3}{12}$ $\dfrac{6}{24} = \dfrac{3}{12}$

Code :   1,2,3,4   1,2,4,5   1,2,4 / 1   1,4 / 2,8

$\bar{a}$: $\dfrac{7}{15}$ $\dfrac{6}{15}$ $\dfrac{5}{15}$ $\dfrac{4}{15}$
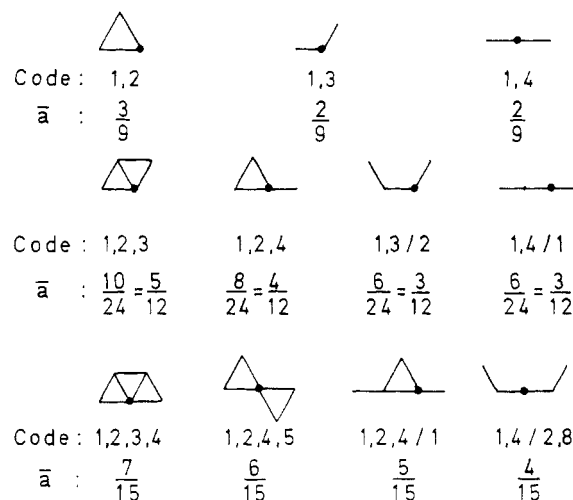
Figure 3. Correlation between focal codes and mean degree of fusion, $\bar{a}$.

Code: 1,4 / 1,7 / 1,2 ,10   1,3 / 1,3,4 / 1,3
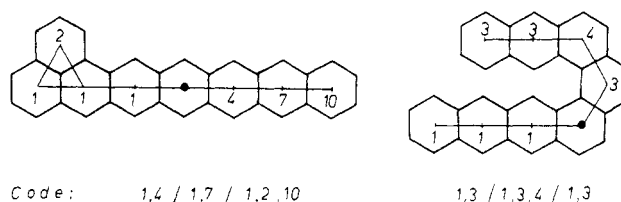
Figure 4. Structure of a perifusene (left-hand formula) whose focal code, on being read as a number, is larger than that of a catafusene with the same number, n, of rings (right-hand formula).

3 seem to confirm this assumption, indicating a general trend for the focal code to decrease (on being read as a number) when the molecular cyclicity increases. In these examples, the mean degree of fusion, $\bar{a}$, of each molecule is given. This quantity was defined[20] by averaging the degrees of fusion, a_i, for each ring of the molecule. In turn, a_i is defined as a fraction whose numerator is the number of sites of fusion for each ring (the number of C–C bonds shared with other rings) denoted by k_i, while the denominator is the number of all edges of the rings (in all polyhexes rings are six membered) denoted by $k = 6$; the sum is averaged by dividing by the number of rings, n.

$$a_i = \frac{k_i}{k} = \frac{k_i}{6} \qquad \bar{a} = \frac{1}{n} \sum_{i=1}^{n} a_i = \frac{1}{6n} \sum_{i=1}^{n} k_i$$

The benzenoid catafusenes have all the same average degree of fusion; in the internal rings the degree of condensation is $2/6$ while in the two terminal rings it is $1/6$ so that the average is

$$\bar{a}_{cata} = [(n-2)2/6 + 2/6]/n = (n-1)/(3n)$$

For perifusenes, $\bar{a}$ will increase, as compared with cata-fusenes, by $1/(3n)$ for each new site of fusion. Denoting the additional sites of fusion in perifusenes by j, one obtains for cata- and perifusenes with the same n

$$\bar{a}_{peri} = (n - 1 + j)/3n = \bar{a}_{cata} + j/3n$$

Since the code decreases usually when the mean degree of fusion increases, the following approximate inequality holds for equal n values; when read as a number, $code_{perifusenes} <$ $code_{catafusenes}$. An exception to this rule are polyhexes with peri fusion in a terminal position of a long acene, whose code leads to a larger number than the code of some catafusenes with the same n (Figure 4).

On examining a set of n-fusenes ordered according to their focal codes, one finds at the end nearly all catafusenes with

larger codes (when read as numbers, ignoring commas and slashes, in a numeration system with a sufficiently larger base than 10) while practically all perifusenes have smaller codes.

The analysis of Tables IV and V provides additonal information for discriminating between cata- and peri-condensed systems by their codes only. Thus, all compounds whose code starts with 1,2 are perifusenes, since these digits indicate the smallest perifusene, namely, the perinaphthenyl radical. Irrespective of which shell it refers to, the digits 1,2 after a slash indicate a perifusene. Also, if the first two digits in any shell are the same as in the preceding shell, the compound is a perifusene. Another rule for molecules with $n \leq 5$ is that only catafusenes have the combination of digits 1,3 in the first shell.

In a previous approach to correlate three-digit codes with molecular properties, it was noted[11] that a biparametric equation using the number n of benzenoid rings and the number of zeroes in the three-digit code gives good agreement with resonance energies calculated by HMO or more refined methods.[26] An analogous correlation can be made with the new code, though less simply, by noting how many times a rectilinear subgraph O—O—O occurs in the dualist graph, similarly to a topological index.[27] In many instances, when the digit 1 appears twice or more times at the beginning of each shell, or when the code has 1,4 in the first shell, this indicates such rectilinear portions of the dualist graph which lead to lower resonance energies.

NOMENCLATURE OF FUSENES

The focal code developed for fusenes may be used as a more convenient device for the nomenclature of any such compound. The code in brackets is used as a prefix for specifying the topology of the fusene. It is followed by the Greek numeral indicating n, then (facultatively) by cata, peri, or corona (whereby peri may include portions of cata condensation, while corona may include portions with peri-, cata-, or peri + cata condensation), and at the end the suffix fusene is added. Though the indication of the n value and the kind of condensation in the name are redundant, we consider this redundancy useful, also with regard to future work which will examine nonbenzenoid systems. As examples, the two carcinogenic hydrocarbons dibenzo[a,h]anthracene and benzo[b]pyrene correspond in the newly proposed nomenclature to the names (1,4/2,8)-pentafusene (pentacatafusene) and (1,2,3/1)-pentafusene (or pentaperifusene), respectively. The simplicity of the proposed nomenclature over the existing IUPAC,[28] Ring Index,[29] or Chemical Abstracts systems[30] is evident. A great advantage of the proposed topological nomenclature resides in its generality (it may be applied to all polyhexes) and in assigning each benzenoid ring a unique number in the code; the latter feature makes it possible to imagine a nomenclature system which will apply also to substituted derivatives, which is not discussed in detail here. In both these aspects, the present nomenclature proposal is superior to the previous one.[1,2]

The fact that there may be found a privileged vertex in any graph (the focal point) on which a coding and nomenclature system may be based gives this approach a much wider scope. We intend to develop the focal coding and nomenclature also for other classes of compounds. Finally, it must be stressed that, owing to its topological nature, the new approach is more

than a simple alternative way of coding and naming molecules; it also contains important structural information which correlates directly with several molecular properties. The existing nomenclature systems are unable to provide such an insight into the dependence of properties upon chemical structure. Studies on these correlations between the new coding system and various properties of benzenoid hydrocarbons are now in progress.

REFERENCES AND NOTES

(1) A. T. Balaban and F. Harary, *Tetrahedron*, **24**, 2505 (1968).
(2) A. T. Balaban, *Tetrahedron*, **25**, 2949 (1969).
(3) A. T. Balaban, *Rev. Roum. Chim.*, **15**, 1251 (1970).
(4) A. T. Balaban, *Rev. Roum. Chim.*, **18**, 635 (1973).
(5) A. T. Balaban and P. von R. Schleyer, *Tetrahedron*, **34**, 3599 (1978).
(6) N. Lozac'h, L. A. Goodson, and W. H. Powell, *Angew. Chem.*, **91**, 951 (1979).
(7) R. C. Read and R. S. Milner, "A new system for the designation of chemical compounds for the purposes of data retrieval. I. Acyclic compounds". R. C. Read, "The coding of trees and tree-like graphs", Preprints, University of the West Indies, Jamaica, 1969; a revised version of the former preprint has been prepared as Research Report CORR 78-42 at the Faculty of Mathematics, University of Waterloo, Canada, 1978.
(8) J. Lederberg, in "The Mathematical Sciences", MIT Press: Cambridge, MA, 1969, p 37; J. Lederberg, G. L. Sutherland, B. G. Buchanan, E. A. Feigenbaum, A. V. Robertson, A. M. Duffield, and C. Djerassi, *J. Am. Chem. Soc.*, **91**, 2973 (1969); J. Lederberg, *Proc. Natl. Acad. Sci. U.S.A.*, **53**, 134 (1965); B. G. Buchanan, A. M. Duffield, and A. V. Robertson, in "Mass Spectrometry. Techniques and Applications", G. W. A. Milne, Ed., Wiley Interscience, New York, 1971, p 121.
(9) F. Diederich and H. A. Staab, *Angew. Chem.*, **90**, 383 (1978); C. Krieger, F. Diederich, D. Schneizer, and H. A. Staab, *ibid.*, **91**, 733 (1979). For theoretical discussions of coronaphenes, see R. McWeeny, *Proc. Phys. Soc. A*, **64**, 261, 921 (1951); G. Ege and H. Fischer, *Tetrahedron*, **23**, 149 (1967); O. E. Polansky and D. H. Rouvray, *Math. Chem.*, **3**, 97 (1977).
(10) R. C. Read and F. Harary, *Proc. Edinburgh Math. Soc.* 2, **17**, 1 (1970).
(11) A. T. Balaban, *Rev. Roum. Chim.*, **15**, 1243 (1970).
(12) K. Balasubramanian, J. J. Kaufman, W. S. Koski, and A. T. Balaban, *J. Comput. Chem.*, **1**, 149 (1980).
(13) D. Bonchev, A. T. Balaban, and O. Mekenjan, *J. Chem. Inf. Comput. Sci.*, **20**, 106 (1980).
(14) M. Randić, *J. Chem. Inf. Comput. Sci.*, **18**, 101 (1978).
(15) M. Randić, *Math. Chem.*, **7**, 5 (1969).
(16) F. Harary, "Graph Theory", Addison-Wesley, Reading, MA, 1972.
(17) R. J. Wilson, "Introduction to Graph Theory", Longman, London, 1972.
(18) D. H. Rouvray and A. T. Balaban, "Applications of Graph Theory", R. J. Wilson and L. W. Beineke, Eds., Academic Press, London, 1979; A. T. Balaban, "Chemical Applications of Graph Theory", A. T. Balaban, Ed., Academic Press, London, 1976.
(19) D. Bonchev, A. T. Balaban, and M. Randić, *Int. J. Quantum Chem.*, **19**, 61 (1981).
(20) D. Bonchev, O. Mekenjan, and N. Trinajstić, *Int. J. Quantum Chem.*, **12**, 293 (1978).
(21) H. Wiener, *J. Am. Chem. Soc.*, **69**, 17, 2636 (1947); *J. Chem. Phys.*, **15**, 766 (1947); *J. Phys. Chem.*, **52**, 425, 1082 (1948).
(22) D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **67**, 4517 (1977).
(23) D. Bonchev, J. V. Knop, and N. Trinajstić, *Math. Chem.*, **6**, 21 (1979).
(24) A. T. Balaban, *Theor. Chim. Acta*, **53**, 355 (1979).
(25) A. T. Balaban, A. Chiriac, I. Motoc, and Z. Simon, "Steric Fit in Quantitative Structure-Activity Relations", *Lect. Notes Chem.*, **15**, 1980.
(26) V. E. Sahini, *J. Chim. Phys.*, **59**, 177 (1962); *Rev. Chim. Acad. Repub. Pop. Roum.*, **7**, 1265 (1962); *Rev. Chim. (Bucharest)*, 551 (1964); V. E. Sahini and A. Savin, *Rev. Roum. Chim.*, **22**, 39 (1977); **24**, 165 (1979).
(27) M. Gordon and G. R. Scantlebury, *Trans. Faraday Soc.*, **60**, 605 (1964).
(28) IUPAC, "Nomenclature Rules for Organic Chemistry", Butterworths, London, 1965.
(29) A. M. Patterson, L. T. Capell, and D. F. Walker, "The Ring Index", 2nd ed., American Chemical Society, Washington, DC, 1960.
(30) Index Guide Introduction, *Chem. Abstr.*, **76** (1972).