

# A Simplified Algorithm Using Base 5 Numbers To Assign Canonical Names to Cata-Condensed Polybenzenes

Seymour B. Elk

Elk Technical Associates, 321 Harris Place, New Milford, New Jersey 07646

Received December 27, 1993\*

Because of the many mathematical problems encountered in the ongoing process of trying to develop a unique and systematically logical (i.e., canonical) nomenclature for chemical molecules, it has been expedient to focus attention on smaller, but chemically important, subclasses. For one such subclass, polycyclic aromatic hydrocarbons consisting of all rings of size 6, many ingenious and highly complex combinations of ideas emanating from various branches of mathematics (especially graph theory, topology, and more recently number theory) have been employed. In this report, we have formulated a simplified algorithm for deriving a canonical name for any assemblage of cata-condensed hexagonal rings (i.e., benzene modules) using base 5 numbers.

Following up on the various nomenclature systems that we<sup>1</sup> have formulated for that limited class of molecules which may be represented by the edge fusion of coplanar regular hexagons (herein referred to as polybenzenes), we now present a greatly simplified algorithm for assigning nomenclature to the further limited class of cata-condensed<sup>2</sup> rings, where cata-condensed means that at no vertex are three hexagons coincident.

Examining the hexagonal tessellation of the plane in standard (Patterson) orientation<sup>3</sup> (Figure 1), we note that for an arbitrarily chosen hexagon (herein denoted with an asterisk), we may rotate this system of hexagons so that any of its six neighbor hexagons is located horizontally to the right of the starred hexagon. Using these two hexagons as a vector, there are exactly three hexagons that are cata-condensed to the second of the described hexagons. To these we assign the following designations (Figure 2):

- 1, if this hexagon continues the chain in the same direction;
- 2, if, in the given planar tessellation, this hexagon is counterclockwise to the starting vector; and
- 3, if, in the given planar tessellation, this hexagon is clockwise to the starting vector.

Note that these three orientations are unfilar connections of the given hexagons.<sup>4</sup>

There exists one additional combination of cata-condensed hexagons that is of importance—when two hexagons can be annellated to the starting vector such that these hexagons are a combination of 2 and 3 above. This is referred to as polyfilar<sup>4</sup> and shall be designated by 4 (Figure 3). Also, in setting up the algorithm, we need to specify when a chain ends. This we designate as 0. Table 1 summarizes this assignment.

Given this assignment table, the only step missing is how to determine where to start. At this point, we note that any hexagon of the molecule having either only one neighbor (i.e., an end hexagon) or else two noncollinear neighbors may be chosen as the starting point (designated by an asterisk). Also, any of the six adjacent hexagons that is a ring of the given molecule will serve as the second ring of the forming vector. For example, in the tetradecebenzene shown in Figure 4, we have arbitrarily chosen that the molecule be centered and that the leftmost hexagon on or near the principal line be the one that is designated as the starting point (starred). This

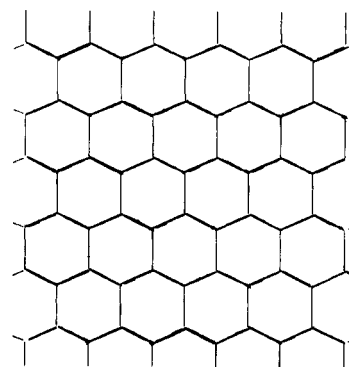


Figure 1.

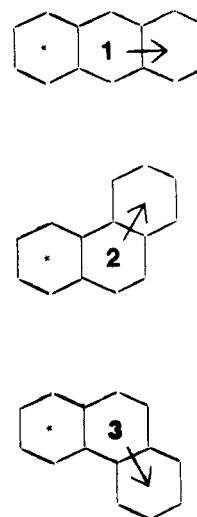


Figure 2.

Table 1. Number Assignment in a Cata-Condensed Polybenzene

0	end of chain
1	extended chain in same direction as the vector
2	extend chain in counterclockwise direction
3	extend chain in clockwise direction
4	extend chain in both counterclockwise and clockwise directions

gives rise to the number tree shown in Figure 5 and a 14-digit name, such as:

4-3-0-4-1-0-1-4-2-0-4-0-1-0

\* Abstract published in *Advance ACS Abstracts*, April 1, 1994.

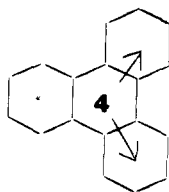


Figure 3.

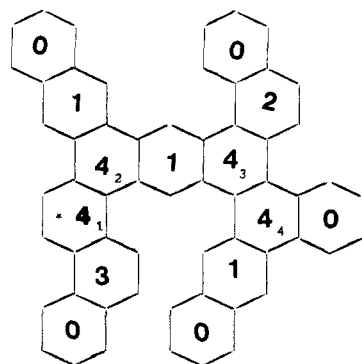


Figure 4.

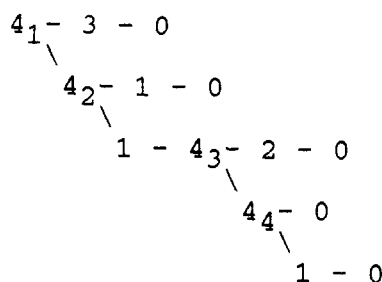


Figure 5.

Before progressing further, however, we note that at every hexagon having two exiting vectors (i.e., assigned the number 4), it was purely random fiat which list to follow first. In other words, we might just as well have named this string:

4-4-1-0-1-4-2-0-4-0-1-0-3-0

or

4-3-0-4-1-0-1-4-4-0-1-0-2-0, etc.

Now, in order to introduce standardization, let us limit the choice of starting hexagon to one of the end rings. A very desirable consequence of this choice is that we now generate a sequence of numbers that is one less than the number of hexagons. The reason for this diminution of identifying digits is that the 0 used to indicate the end ring is not needed for the starting hexagon. Furthermore, since we completely cover any such assemblage of hexagons terminating at a different end hexagon, the last digit of the name must also be a 0, which can also be omitted without ambiguity; consequently, we can assign a name that is  $h - 2$  digits long (where  $h$  is the number of hexagons in the molecule). Figure 6 (rotation of the molecule shown in Figure 4 through  $+60^\circ$ ) similarly produces a different numbering of the hexagons, from which we can form a second number tree (Figure 7). The similarity of these two number trees (Figures 5 vs 7) is evident, with the only differences being the omission of a 0 at the beginning of the Figure 6 sequence and a different digit corresponding to the starting hexagon of Figure 4.

The final step is to note that if we have  $e$  end hexagons, we could have started with any one of them and ended with any

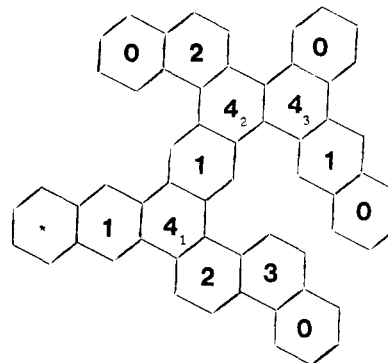


Figure 6.

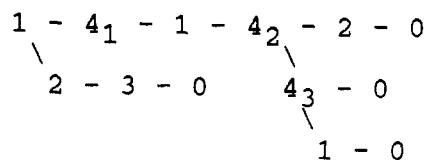


Figure 7.

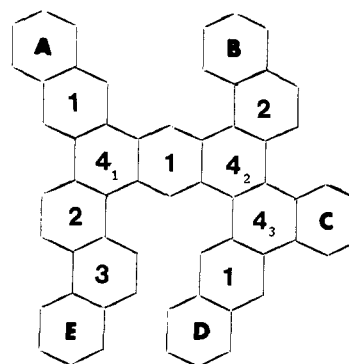


Figure 8.

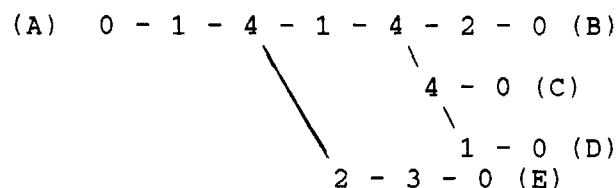

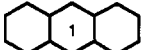
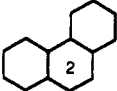
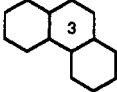
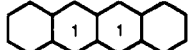
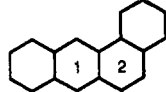
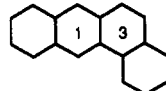
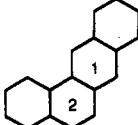
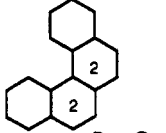
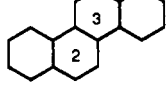
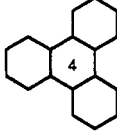


Figure 9.

of the other  $e - 1$  end hexagons. In other words, at first glance, it appears that we must examine a minimum of  $e(e - 1)$  combinations of starting and ending hexagons—actually more since there can exist several options for each, depending on how many possible pathways may be traversed. The actual number of such pathways can be further determined by the observation that the number of end rings is always exactly two more than the number of polyfilar rings. Consequently, the number of paths to examine for a cata-condensed polybenzene, herein designated as  $P(e)$ , is  $e2^{e-2}$ , i.e.,  $P(2) = 2$ ,  $P(3) = 6$ ,  $P(4) = 16$ ,  $P(5) = 40$ , etc. From this number of possible pathways, we now wish to select the “optimal” one, according to some preselected criterion. The simplest (which is also the intuitive) choice is to pick the lowest base 5 number (since each digit can range from 0 to 4) from this set of  $h - 2$  digits.

In order to implement this process, let us randomly label the five end hexagons A–E (see, for example, Figure 8 and the corresponding tree in Figure 9). Next, we wish to tabulate trees with starting and ending hexagons AB, AC, ..., ED. However, rather than the brute force tabulation of each of these 40 cases, we note the simplification that once we have

Table 2. Correlation of Sequential Code with Molecules

	code	molecule	
dibenzene	00		
tribenzenes	010 = 1		
	020 = 2		
	030 = 3		same as 2
	040 = 4	a terminal 4 in the abbreviated code is impossible	
tetrabenzenes	0100 = 10	IMPOSSIBLE: two zeroes in sequence can occur ONLY following a polyfilar hexagon, i.e., following a 4 in the code	
	0110 = 11		
	0120 = 12		
	0130 = 13		
	0140 = 14	a terminal 4 in the abbreviated code is impossible	
	0200 = 20	IMPOSSIBLE: same reason as for 0100	
	0210 = 21		same as 12
	0220 = 22		
	0230 = 23		
	0240 = 24	a terminal 4 in the abbreviated code is impossible	
	03xx	all combinations starting with 03 are the mirror images of a code starting with 02	
	0400 = 40		
	04x0 = 4x (x = 1, 2, 3, or 4)	IMPOSSIBLE: a 4 in the code describes two succeeding hexagons; thus in the abbreviated code, the presence of a 4 followed by a 1, 2, or 3 would imply five hexagons in this system; similarly, a second 4 would need six hexagons	

chosen a numbering algorithm, we can ignore certain selected paths, as they will give a higher base 5 number. For example, the code for AB must include following the path to E first when encountering the first polyfilar hexagon (labeled with a 4) and also must follow the path to C and D first when encountering the second polyfilar hexagon—otherwise, it would not be possible to have a path ending at B. On the other hand, when encountering the third polyfilar hexagon, the choice of ordering would be completely arbitrary were it

not for the numbering algorithm. The net result of this is that, disregarding for the moment such a numbering algorithm, possible 12-digit names (omitting the end zeroes) for AB are

1-4-2-3-0-1-4-4-0-1-0-2

or

1-4-2-3-0-1-4-4-1-0-2

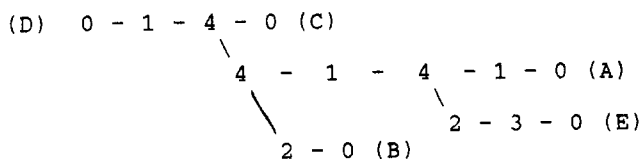
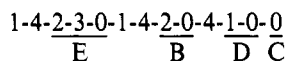


Figure 10.

Similarly, in developing the name for AC, we note that the path is completely unambiguous, going first to E, then to B, next to D, and finally to C:



where the final 0 can be dropped. Conversely, the path for AE can go through: (1) first B, then C, then D; (2) first B, then D, then C; (3) first C, then D, then B; or (4) first D, then C, then B. Rather than considering all mathematically possible combinations, it is pragmatic to introduce our algorithm for assigning nomenclature early, i.e., at every polyfilar junction encountered when traversing a path through all of the rings, we shall choose the lowest base 5 number as the desired name. We thus see that the choice of assignments for 1, 2, 3, and 4 made earlier is as would be necessary to fulfill Patterson's objectives, namely, the lowest nonzero number elongates the chain along the principal axis. This is shown in the example by beginning with either A or D since these two have three hexagons in the longest chain, rather than B, C, or E which have only two. Next, we prefer hexagons that lie above the principal axis to those below; consequently, counterclockwise orientation is given precedence over clockwise. Finally, terminating the chain is the simplest operation and so is given highest priority (assigned the number 0), while branching is the most complicated of the annellations and so is considered last (assigned the number 4).

Finally, we note that the tree given in Figure 9 could have been drawn starting from D and following the lower numbered branch first whenever encountering a 4 in the code (see Figure 10). In other words, the lowest number we can form is the canonical name. In this example, it was derived by starting at D, progressing to C, then to A, then to E, and finally to B. Consequently, the 12-digit canonical name is 1-4-0-4-1-4-1-0-2-3-0-2.

Table 2 contains the correlation of code with each of the possible cata-condensed di-, tri-, and tetrabenzenes. Note that the first entry of this table, dibenzene (naphthalene), is simply the vector itself, as both the starting and final zeros in the name have been omitted. Table 3 lists the canonical names and the formation of the 12 possible cata-condensed pentabenzenes.

Table 3. Canonical Names for Cata-Condensed Pentabenzenes

111		112	
121		122	
123		140	
212		213	
222		223	
232		240	

## REFERENCES AND NOTES

- (1) (a) Elk, S. B. A Nomenclature for Regular tessellations and its Application to Polycyclic Aromatic Hydrocarbons. *MATCH* **1980**, *8*, 121-158. (b) Elk, S. B., Formulation of a Canonical Nomenclature for Polybenzenes Using Triangular-shaped Hexagonal Tessellation Envelopes. *MATCH* **1985**, *17*, 255-68. (c) Elk, S. B. Canonical Orderings of Hexagons That: (1) Nearly Meet the Intent of Patterson's Nomenclature Rules and (2) Orient a Molecule for Purposes of Assigning a IUPAC Name. *Polycyclic Aromat. Compd.* **1990**, *1*, 109-21. (d) Elk, S. B. A Canonical Ordering of Polybenzenes and Polymantanes Using a Prime Factorization Technique. *J. Math. Chem.* **1990**, *4*, 55-68.
- (2) (a) Balaban, A. T.; Harary, F. Chemical Graphs 5. Enumeration and Proposed Nomenclature Cata-condensed Polycyclic Aromatic Hydrocarbons. *Tetrahedron* **1968**, *24*, 2505-14. (b) Balaban, A. T. Chemical Graphs 7. Proposed Nomenclature of Branched Cata-condensed Benzenoid Polycyclic Hydrocarbons. *Tetrahedron* **1969**, *25*, 2949-56.
- (3) Patterson, A. M. Proposed International Rules for Numbering Organic Ring Systems. *J. Am. Chem. Soc.* **1925**, *47*, 543-61.
- (4) Taylor, F. L. Enumerative Nomenclature for Organic Ring Systems. *Ind. Eng. Chem.* **1948**, *40*, 734-8.