

A NOMENCLATURE FOR REGULAR TESSELLATIONS
AND ITS APPLICATION TO POLYCYCLIC AROMATIC HYDROCARBONS

Seymour B. Elk

New Jersey Institute of Technology

(received: February 1980)

In order to develop a unique and systematically logical (i.e., canonical) nomenclature for polycyclic aromatic hydrocarbons, techniques from geometry and graph theory are applied to the regular hexagonal tessellation of the plane. Unlike the IUPAC¹ nomenclature which starts with thirty-five "common names" and adds benzene rings onto previously named compounds in a mathematically uncoordinated manner, algorithms of nomenclature schemes for connected chains of benzene rings are presented in Part I and a simplified scheme for large clusters in Part II. For the chains of benzene rings, a synthetic algorithm -- which is fairly close to the IUPAC method of naming alkanes -- is first presented. An analytic algorithm, which requires no preferential orientation (and thus can be done by a computer), is next presented. For cluster molecules the topological problems of multiply-connected regions make these algorithms cumbersome; therefore, a simplified scheme is presented.

In addition, a unifying concept, the characteristic number, is developed for the various configurations possible with hexagonal tessellations. Also, the use of the other planar tessellations is examined -- including the problems associated with their usage.

1. INTRODUCTION

In 1892 an international convention proposed the first systematic nomenclature for organic compounds. This has been continually up-graded to include the names for new chemicals as they are discovered, under the auspices of the International Union of Pure and Applied Chemistry, IUPAC¹. Despite the fact that IUPAC has employed a large amount of geometric reasoning (due to the geometrical nature of the compounds being named), no purely geometrically based system has been presented. This has resulted in the proliferation of numerous common names for the skeletal structure of the class of polycyclic aromatic hydrocarbons, as well as ad hoc addition schemes for naming new polycyclics, with the concomitant result that many of the geometric, topologic and graph theoretic properties of these compounds have been overlooked.

This paper describes the simple geometry of the three regular planar tessellations, with particular emphasis on the hexagon tessellation -- which is then used as the basis for forming systematic nomenclature schemes for polycyclic aromatic hydrocarbons. Polycyclic aromatic hydrocarbons are perfectly modelled by the hexagonal tessellation since these compounds are strictly planar and contain only a single element (carbon) that must be geometrically positioned. The presence of the other element (hydrogen) is determined solely by the geometry of the carbons and the simple need for bond saturation. In other words, from a mathematical perspective, the system reduces to specifying only one parameter. Notwithstanding this geometrical "simplicity", the development of such a system of nomenclature has not been previously undertaken along the lines proposed.

In addition, a superficial examination of the regular square tessellation and its relation to the systematic nomenclature of the alkanes and of the regular triangular tessellation as the basis for naming hexa-coordinated compounds, such as the author described in Reference 2 (if and when a sufficient number of such compounds are isolated in order to make the task of interest to the practicing chemist) are presented.

I. CHAIN MOLECULES

2. TESSELLATIONS

If an entire space, such as a line, plane or three-dimensional space, is subdivided into an infinite array of connected congruent figures, the space is called "tessellated" and the union of these figures is a "tessellation". Although artistically there exist an infinite variety of such figures, the tessellations of interest to chemists are unions of regular polygons or polyhedra.

Because the concept of shape is not relevant to line segments, the simplest geometrically interesting tessellations occur in a plane and are formed using congruent regular polygons: Beginning with the regular (equilateral) triangle, six such triangles may be joined without overlapping at a single point. (Figure 1). Each of the other vertices of these tri-



Figure 1

angles may again be considered the center for six triangles (Figure 2), etc. In this manner the entire plane may be covered by the union of connected congruent triangles. Similarly,

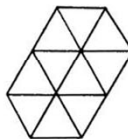


Figure 2

beginning with the next simplest regular

polygon (the square), the plane is tessellated by four squares

joined at each vertex (Figure 3). Continuing to the regular pentagon, note that each internal angle is 108° . Therefore, three regular pentagons cover only 324° while four regular pentagons would overlap; consequently, no

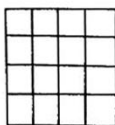


Figure 3 tessellation involving reg-

ular pentagons exists. The regular hexagon, on the other hand, has an internal angle of 120° and thus forms the third of the regular tessellations (Figure 4). For larger regular polygons each internal angle is greater than 120° but less than 180° ; therefore, they can not form tessellations.

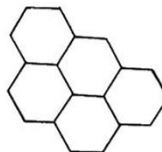


Figure 4

In three dimensions, of the five regular polyhedra, only

the cube tessellates. Because each point of the cubic tessellation is the union of six edges, pictorially there is a degree of similarity to the triangular tessellation, especially for small molecules -- the only ones capable of being dealt with efficiently.

3. HEXAGONAL TESSELLATIONS AND POLYCYCLIC AROMATIC HYDROCARBONS

Of the three regular planar tessellations, the simplest (in terms of minimizing the number of line segments meeting at a vertex), the hexagon has traditionally been the basis for describing aromatic compounds. This is true since each benzene ring contains six equally spaced carbon atoms and is essentially planar. In other words, the geometric representation does not introduce distortion (as does the square tessellation when applied to alkanes -- see Section 8).

The IUPAC'S nomenclature scheme¹ lists 35 polycyclic aromatics with "common names" and adds benzene rings onto these previously named compounds in a mathematically uncoordinated manner in order to build up the basic skeletal structure.³ Attempts at standardization begin for the derivatives of these compounds, rather than going back to the basic structure itself.

By limiting attention to only six member rings lying in a plane, nomenclature schemes are developed that are fairly straightforward and are closely akin to the systematic nomenclature of the alkanes. Two types of algorithms are presented: Algorithms 1 and 2 are intended for connected chains of benzene rings and are described below. The second type, which will be described in Part II, is especially suited to large clusters of benzene rings. Note that none of these nomenclature schemes are relevant to non-co-planar systems such as Dewar "benzene" or helicene.

Beginning from a single benzene ring, there are six edges to which a second benzene ring may be affixed, as well as six vertices. Because of the symmetry of the benzene molecule, each edge is identical to every other edge and, although there exist six permutations, there exists only one combination -- the naphthalene molecule (Figure 5).

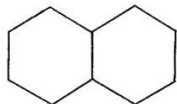


Figure 5

Examining next two benzene rings joined at a common vertex (Figure 6): Such a configuration would imply either six bonds at the common carbon atom in violation of all known chemical



Figure 6

phenomenon, or else four tetrahedrally oriented bonds -- non-co-planar. Therefore, no further study of such an array is required and all fusions between rings will occur at a common edge or series of edges. Because of the uniqueness of the naphthalene molecule, the simple name "dibenzene" will suffice.

Although there exist ten edges at which it appears possible to add a third ring to dibenzene, two sets (labelled 3 & 3' and 7 and 7' in Figure 7) are identical -- a ring added at position 3 will contain edge 3' and vice versa (due to the tessellating nature of the hexagon; similarly, for edges 7 and 7'). Thus there exist eight distinguishable edges. However, the rings themselves are indistinguishable; therefore, adding a ring at position 1 is equivalent to adding one at position 5; similarly, for positions 2, 4, 6 and 8, and for positions 3 and 7. In other words, because of symmetry considerations, there exist only three possible different positions to which a benzene ring may be affixed.

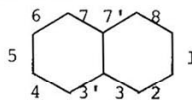


Figure 7

From a mathematical perspective all three of these forms are achievable. These may be named: a-tribenzene, b-tribenzene and c-tribenzene (Figure 8). From a chemical perspective,

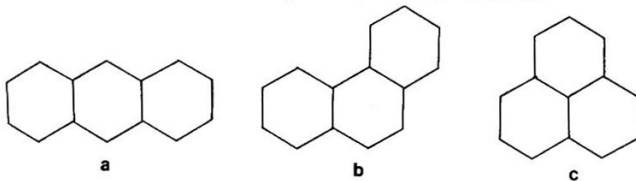


Figure 8

a-tribenzene and b-tribenzene (called "anthracene" and "phenanthrene" respectively) have formula $C_{14}H_{10}$ and their existence and properties are well established. c-Tribenzene, on the other hand, has formula $C_{13}H_9$ and has not been found to exist. That there exists problems with c-tribenzene may be seen by noting that each carbon atom has four bonds and each hydrogen one. Since, in its simplest form a bond exists between two atoms and there are an odd number of bonds (61), one of the carbons must have only three bonds, such as shown in Figure 9

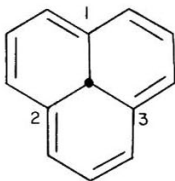


Figure 9

at the point of union of the three benzene rings. Such a point may be referred to as a "triple point". If the traditional Kekule structure* is allocated to this molecule, a conjugated system of single and double bonds can not be accommodated without there existing a single unpaired electron (or else an extra hydrogen atom is required). See Figure 10. The presence of the hydrogen at the indicated spot yields the compound phenalene, which is not a pure aromatic. The alternate Kekule structure, shown in Figure 9, suggests the possibility of forming a sandwich dimer $C_{26}H_{18}$. However, no such compounds of this type are presently known.

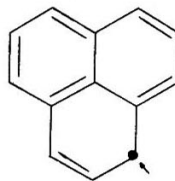


Figure 10

Returning to the mathematical perspective, note that there exist three double vertices (labelled 1, 2 and 3 in Figure 9) and one triple vertex. The presence of an odd number of triple vertices is always accompanied by an odd number of hydrogen atoms or else a break in the conjugation. A hydrocarbon in which there is a demand for both conjugation and an odd number of hydrogen atoms has not been found and is assumed to be "chemically unattainable". The logical extension to the sum of nine single vertices plus one triple vertex (thus an even number of odd vertices) lends credence to the sandwich dimer

*The existence of a Kekule structure is not a necessity.

hypothesis -- in that a path exists but it does not end at the starting point.⁴

Unlike the technique employed in the naming of alkanes, start by examining the synthesis process used to produce polycyclic aromatic hydrocarbons. In other words, continue the process used to determine that there exist only three mathematically distinct tribenzenes; namely, for each of these tribenzenes affix a hexagon at every conceivable location. From a combinatorics perspective, at each of the ten edges of a- and b-tribenzene and at each of the nine edges of c-tribenzene examine the new molecule formed by affixing a single ring. From symmetry considerations, however, this is extremely wasteful -- only three locations for a-tribenzene, six locations for b-tribenzene and two locations for c-tribenzene need to be examined. All the remainder are merely duplications.

Relying on the rules established for orientation of ortho and peri systems by the IUPAC¹, a unique nearly consistent naming of these eleven tetrabenzenes is set up that is easily expandable to larger systems; however, in the process, a liberal amount of physical intuition is employed -- Algorithm #1. See Appendix.

Using capital letters for this synthesis approach, the three tetrabenzenes derived from A (=a)-tribenzene are named AA-, BA- and CA-tetrabenzene respectively. See Table 1. Next, examining the six tetrabenzenes derived from B-tribenzene, two duplications are noted: AB- is the same molecule as was BA- and EB- is the same as CA-tetrabenzene. Because it is expedient, rather than necessary, to eliminate redundancy, choose the lexicographically lowest name starting from the right and working to the left. In other words, BA- and CA-tetrabenzene are the preferred names. Similarly, each tetrabenzene formed starting from C-tribenzene has been previously named.

Although the synthesis method is a "viable" scheme for naming small molecules without the need for a computer, it may require a tedious series of frequent reorientations of the molecule with concomitant recalculations after each ring has been added. As the molecule gets larger the criterion for orienta-

Derived from	Structure	Synthetic Name	Configuration	Interim Name	Analytic Name
B C C B		AA-		aa-	aa-
A * * * A				fa-	
B C C B					
A-Tribenzene					
		BA-		$\begin{matrix} (a_1) \\ (a_1) \end{matrix}$	$\begin{matrix} e_1 a_1 \\ g_2 a_1 \end{matrix}$
				$\begin{matrix} (a_2) \\ (a_2) \end{matrix}$	$\begin{matrix} j_1 a_2 \\ b_2 g_2 \end{matrix}$
		CA-		$\begin{matrix} (a_1) \\ (a_1) \\ (a_2) \end{matrix}$	$\begin{matrix} d_1 a_1 \\ h_2 a_1 \\ i_1 a_2 \\ c_2 a_2 \end{matrix}$
				$\begin{matrix} (a_2) \end{matrix}$	

TABLE 1 (Page 2)
TETRABENZENES

Derived From	Structure	Synthetic Name	Configuration	Interim Name	Analytic Name
B A B C * D A * * E D E F B-Tribenzene		BB-		d ₁ d ₁	db-
				h ₂ f ₂	
				h ₁ h ₁	
				d ₂ b ₂	
				f ₁ b ₁	
				j ₂ f ₂	

By symmetry,
a₄=a₃; a₅=a₁

TABLE 1 (Page 3)
TETRABENZENES

Derived From	Structure	Synthetic Name	Configuration	Interim Name	Analytic Name
B-Tribenzene		CB-		$\begin{matrix} e_1 c_1 \\ i_2 f_2 \end{matrix}$	} eb-
	By symmetry, $a_1=a_4=a_5=a_8; a_2=a_3=a_6=a_7$			$\begin{matrix} i_1 g_1 \\ e_2 b_2 \end{matrix}$	
		DB-		$\begin{matrix} f_1 d_1 \\ f_2 f_2 \end{matrix}$	} bb-
				$\begin{matrix} j_1 h_1 \\ b_2 b_2 \end{matrix}$	
	By symmetry, $a_4=a_1; a_3=a_2; a_6=a_5$			$\begin{matrix} h_1 b_1 \\ j_2 d_2 \end{matrix}$	

TABLE 1. (Page 4)

TETRABENZENES

Derived From	Structure	Synthetic Name	Configuration	Interim Name	Analytic Name
B-Tribenzene		FB-		h, d j, b	j, b

By symmetry,
 $a_1 = a_4 = a_6$; $a_2 = a_3 = a_5$

tion becomes confusing and even contradictory, with the result that the above scheme leaves a great deal to be desired. Furthermore, it is very cumbersome for analyzing a given molecule -- exactly the problem of concern to the practicing chemist.

An alternate method that is useful in analyzing a given molecule is described in Algorithm #2. See Appendix. This will be designated by lower case letters. Table I contains both the synthetic and analytic names for the tetrabenzenes and indicates how each of the analytic names was formulated: Notice that each possible location for a was first marked. Then for each of these locations the two possible locations for b (clockwise or counterclockwise) was examined and an interim name allocated. The lexicographically lowest -- reading right to left -- interim name is the analytic name.

The technique for determining the analytic name may be greatly simplified by observing which is the nearest unnamed ring for the various locations of a. For instance, in BA-tetrabenzene, a_1 is far removed from the nearest unnamed ring, while a_2 is adjacent to an unnamed ring -- provided the counterclockwise direction of walk is chosen. Thus by inspection a_2 will yield the lexicographically lowest name -- which becomes the analytic name. In Table II for the development of the pentabenzenes, the table has been shortened to include only those configurations having the lowest possible letter in the far right position.

Two important observations that may be made at this time are:

1. The maximum number of distinguishable pentabenzenes is obtained from the combinatorics of hexagonal polyominoes^{5,6}; however, the actual number achieved is diminished by those configurations having an odd number of triple vertices. Of the seven possible tetrabenzenes, five have no triple vertices (i.e., $C_{18}H_{12}$), one has one triple vertex (i.e., $C_{17}H_{11}$, which has not been chemically achieved) and one ($C_{16}H_{10}$) has two triple vertices. This compound, cb-tetrabenzene (common name

TABLE II
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
B C D C B A * * * * A B C D C B AA- Tetrabenzene		AAA-		aaa -
		BAA-		baa -
		CAA-		caa -
		DAA-		daa -

TABLE II (Page 2)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
E, F B, C, D, *G A, * * * H L, K, J, I		BBA-		gba-
BA- Tetrabenzene		CBA-		fb-a-
		DBA-		eba-
		EBA-		dba-

TABLE II (Page 3)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
BA-Tetrabenzene		FBA-		cba-
		GBA-		bba-
		IBA-		lba-

TABLE II (Page 4)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
BA-Tetrabenzene		JBA-		kba-
		KBA-		jba-
		IBA-		iba-

TABLE II (Page 5)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
D E B C * F A * * * G K J I H CA-Tetrabenzene		OCA-		eca-
		DCA-		dca-
		ICA-		jca-

TABLE II (Page 6)
PENTABENZENES

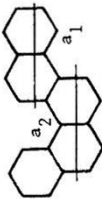
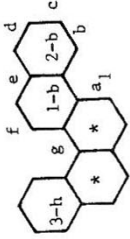
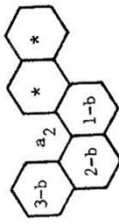
Derived From	Structure	Synthetic Name	Configuration	Analytic Name
CA-Tetrabenzene		JCA-		

TABLE II (Page 7)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
A D B * E B C * F A * * G D E F				
BB-Tetrabenzene				fdb
				ffb
		BBB-		hdb

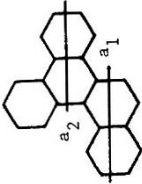
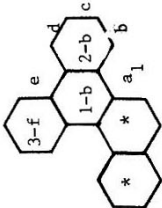
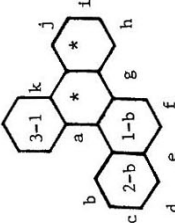


TABLE II (Page 8)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
BB-Tetrabenzene		DBB-		hbb
				bbb

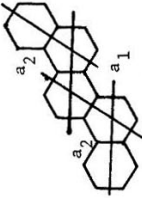
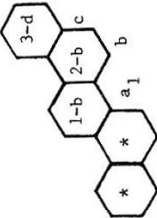
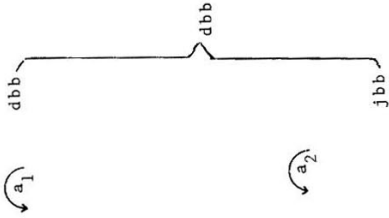
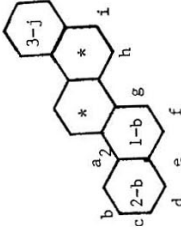
Note that rotating this structure through $+60^\circ$ will give a name whose middle letter is higher than b and thus could not be lexicographically lower than the configurations shown. Similarly, rotating this structure through -60° will give a lexicographically higher name.

TABLE II (Page 9)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
BB-Tetrabenzene		FBB-		fbb
				lbb

See Note on Page 8 of this Table.

TABLE II (Page 10)
PENTABENZENES

Derived From	Structure	Synthetic Name	Configuration	Analytic Name
D E F B C * * A A * * C B F E D		FDB--		
DB-tetrabenzene				

pyrene) both exists and is aromatic.

2. Much of this nomenclature (both schemes) has the highly desirable property of commutativity of symbols; e.g., ba- (or BA-) is equivalent to ab- (or AB-) tetrabenzene; similarly, for ca- and for cb-tetrabenzene. Nevertheless, without some highly unnatural nomenclature scheme, commutativity can not be guaranteed.

Turning now to each of the seven mathematically possible tetrabenzenes: From the 45 combinations generated by adding a benzene ring to each of the different edges of the respective tetrabenzene (4 for AA-, 12 for BA-, etc.) twenty-two different pentabenzene are formed. These are tabulated in Table II, along with both their synthetic and analytic names. Of these twenty-two mathematically possible forms twelve have no triple vertices (i.e., $C_{22}H_{14}$) and three have two triple vertices (i.e., $C_{20}H_{12}$). These fifteen are chemically achievable. The remaining six having one triple vertex (i.e., $C_{21}H_{13}$) and the one having three triple vertices (i.e., $C_{19}H_{11}$ -- ECA-pentabenzene) are chemically unattainable.

At this point it should be noted that with respect to Algorithm #1 some combination of hexagons exceed the simple description used so far:

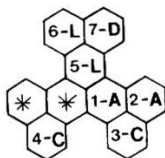


Figure 11

In the molecule shown in Figure 11 each of the three four ring long chains could have been selected as the horizontal -- due to symmetry. This row is named with *, *, 1-A, 2-A. The next ring to be added is obtained by either rotating the molecule 180° thereby yielding the ring in Figure 11 presently below the rings marked 1-A and 2-A, or else by reflecting the molecule horizontally yielding the ring below the two asterisks. In either case (due to symmetry) the selected ring will be designated 3-C. At this point the molecule is top heavy to the upper left and so must be reflected through a vertical line in order to be in canonical form. The next ring to be affixed is then also in position C and thus the choice of

reflecting or rotating is immaterial as the six ring component has the designation CCAA- in either case. The seventh and eighth rings added are below the base line and thus numbered clockwise. These are designated 5-L and 6-L respectively. Now, however, the molecule is bottom heavy and so must be reflected horizontally. The ninth ring is thus named 7-D; thereby, violating the monotonic scheme that was the natural consequence of the algorithm up until this point. By algorithm #1 the molecule is thus named: DLLCCAA-nonabenzene.

Considering next rigidizing the molecule and examining all possible orientations. The analytic name is thus: ollfcaa-nonabenzene. Note that this name has retained the reverse lexicographic ordering of the letters.

The problem of using algorithm #1 becomes even more acute for the combination of hexagons shown in Figure 12: The first

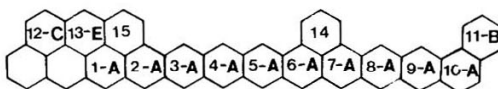


Figure 12

thirteen letters of the name were obtained in a straight-forward manner -- the twelve linear rings were each named with an A (after the base two); next the ring marked 11-B was selected, as this system of 13 rings was lexicographically lowest. The molecule was then reoriented and rings were added at positions C and then E respectively. This structure, however, required reorientation (back to the original, where ring 11-B was on the left) and the fourteenth prefix thus became H and the fifteenth F (thus the name FHECBAAAAAAAAA-heptadecabenzene, or in simpler form FHECB,10-A-heptadecabenzene) -- violating monotonicity.

Even greater confusion arises for the molecule which has a ring (#16) positioned adjacent to rings 8-A and 9-A and with ring 17 omitted. Such a molecule has linear "additive" balance; thereby complicating the choice of orientation. In fact, if the thirteenth ring (11-B) is chosen as the upper right end -- as in Figure 13(a) -- the sixteenth ring is named 14-F, while if ring 12-C is near the right -- as in Figure 13(b) -- this

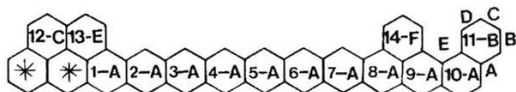


Figure 13(a)

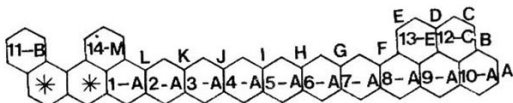


Figure 13(b)

same ring is named 14-M. Furthermore, starting from the hexadecabenzene shown in Figure 13(a) and adding a seventeenth ring there are two different molecules that could be called: MFECB,10-A-heptadecabenzene -- depending on whether the orientation should be shifted (Figure 14(a)) or remain the same

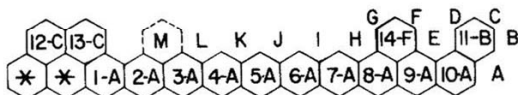


Figure 14(a)

(Figure 14(b)). In fact, determination which is the desired

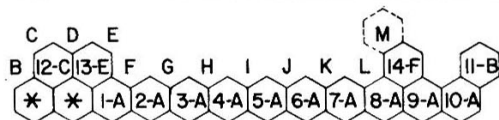


Figure 14(b)

molecule can not be determined at a glance; rather a formal center of gravity integration must be performed. The analytic names for the molecules shown in Figures 12 and 13 are: qpnhb,10a-heptadecabenzene and qofb,10a-hexabenzene.

Examining now two important classes of linear type molecules named in the recent literature^{3,7} as acenes and aphenes:

1. The group of polycyclic aromatics containing anthracene (a-tribenzene), tetracene (aa-tetrabenzene), pentacene (aaa-pentabenzene), etc. is called "acenes". The analytic name for such compounds is $(n-2)a-n$ -benzene; e.g., 4a-hexabenzene. Conversely, any name having only a's is an acene -- with the prefix being two more than the number of rings. Note that naphthalene (dibenzene) could also be included in this naming scheme since $n - 2 = 0$ and no letters were required in order to distinguish between the types of dibenzene -- only one.
2. The group of singly-branched nearly-equal length linear polycyclic aromatics has most recently been called "aphenes". It contains phenanthrene (b-tribenzene), tetraphene (ba-tetrabenzene), pentaphene (cba-pentabenzene), hexaphene (cbaa-hexabenzene), etc. The members of this set are formed by alternately adding a ring to the horizontal and then to the diagonal right end of the previously formed angular system. Although there has been a major improvement by the use of aphene names instead of the older IUPAC names of Benzo-(a)-anthracene and Naphtho-(2'.3':1.2)-anthracene for the four and five ring compounds respectively; the aphene names arbitrarily exclude chains of differing length. For example, without the proposed nomenclature, it is not easy to recognize the familial relation of baa-pentabenzene (not an aphene, IUPAC name benzo-(a)-naphthalene) and cba-pentabenzene (pentaphene). In addition, there exists a special relationship between the prefixed letters that characterize aphenes: The first letter of the analytic name will be -- for an even number of rings the $n/2$ -th letter of the alphabet and for an odd number of rings the $(n+1)/2$ -th letter. Continue down the alphabet for the succeeding rings in the diagonal chain, up to the horizontal chain which are all named a -- excluding the first two. For example, decaphene, undecaphene and dodecaphene (Figure 15) have the names: edcb,4a-decabenzene; fedcb,4a-undecbenzene and fedcb,5a-dodecabenzene respectively.

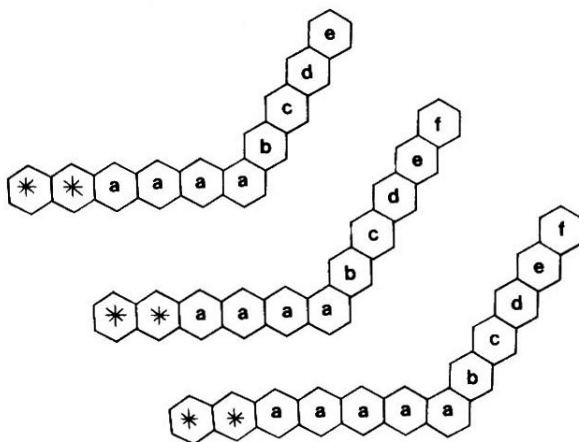


Figure 15

In a similar manner, for any other grouping of polycyclics that has been given a special name, such as the doubly-branched nearly equal length linear polycyclic aromatics referred to as "starphenes"⁷ or the zee-shaped "zethrenes"^{3,7}, these can similarly be systematically named and the symmetries noted so that a comparable repetitive scheme be evolved for the letter prefixes.

II. CLUSTER MOLECULES

4. RINGS OF BENZENE RINGS

In the process of synthesizing polycyclics an important class of compounds, often referred to as "super-aromatics", is encountered. The first example of this is hfdb-hexabenzene (4B-hexabenzene) -- whose common name is "coronene" (Figure 16).



Figure 16

Coronene is the prototype of compounds in which additional rings have been created "interior" to the structure. Geometrically, a five ring compound (fdb-pentabenzene) plus one ring yields a "7-ring" structure. "Topologically" we may view this as a "multiply-connected" (i.e., donut-shaped) six ring structure. Although chemists have traditionally described coronene using the topological scheme, the question is mostly academic inasmuch as the significant chemical properties are unaltered and the "high concentration" electron "pathway" is around any contiguous combination of the 24 edges. Since each edge of the "internal" hexagon is also an edge of one of the "peripheral" hexagons, the question does not arise.

5. CLUSTERED RINGS

The naming of larger compounds of the type described in Section 4, such as Figure 17, leaves leeway for ambiguity that can be easily circumvented by a different type of nomenclature in which clustering is given paramount importance. Unlike the linear chains that have been named so far in which the longest chain is the reference, for clusters (the peri- and ortho- compounds of the

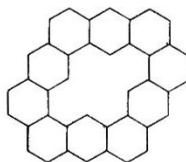


Figure 17

IUPAC nomenclature¹) the new over-riding consideration will be horizontal and vertical symmetry. For example, perylene, (ica-pentabenzene) as a cluster could be written as shown in Figure 18. Note that configuration b is obtained by rotating a through 60° and c by rotating b through 60°. Of these three

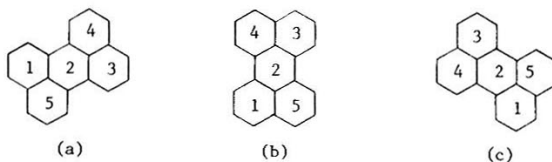


Figure 18

forms, b has the maximum horizontal and vertical symmetry and is selected as the basis for naming clusters -- namely, from top to bottom, read off the number of hexagons. Because of the offset nature of stacked hexagons, in contradistinction to squares, the numbers in this sequence (cited in numerical order when there is a choice) will have to vary by one if there is to be no ambiguity. Of the three configurations in Figure 18 only b meets this requirement; therefore, the name for perylene is: 2,1,2-pentabenzocluster. Other simple clusters formed so far include: eb-tetrabenzene = 1,2,1-tetrabenzocluster; eca-pentabenzene = 2,3-pentabenzocluster; and gdca-hexabenzene = 1,2,3-hexabenzocluster. This latter compound is also referred to by its common name, "triangulene". In addition, coronene (as a seven member ring) would be named 2,3,2-heptabenzocluster.

In order to adapt this nomenclature to a much wider class, x's are added to denote the absence of a ring. In this way each number in the sequence (ring or x's) may be made exactly one more or one less than the preceding entry. For example, perylene in Figure 18a would be: x-1,3,1-x--pentabenzocluster. Similarly, coronene (as the multiply-connected six ring compound) is: 2,1-x-1,2--hexabenzocluster.

Whenever there is the option of choosing a form possessing no x's versus a form with x's, the algorithm shall be to select the form containing no x's -- as this shall always be the one with maximum symmetry. Note that 2,1,2-pentabenzocluster is the preferred name for Figure 18.

The simplicity and universality of this system is seen by examining some sample compounds: Figure 19 (jcaa-hexaben-

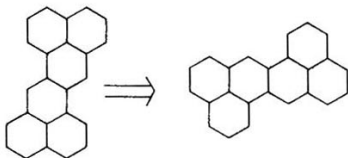


Figure 19

zene) is named: 2x-1,4,1-2x--hexabenzocluster; Figure 20 (feba-hexabenzene) is named: 3,3-x--hexabenzocluster; and Figure 21 (IUPAC name



Figure 20

isoviolan-

threne) is named: 2x-3,x-3-2x,3-4x--nonaben-



Figure 21

zocluster. Notice that for this last compound the x's are needed in order to correctly position the rows. Even with these x's the geometric picture is much more readily obtained than by positioning the seven letter sub-

scripts (**kiddbba-nonabenzene**) of the analytic nomenclature, and is much simpler than the 1.2,8.9 Dibenzopyropyrene of the IUPAC nomenclature¹.

In a similar manner the cluster notation has a decided advantage over the

circum (or circo) prefix with a common name, such as: circumanthracene (Figure 22). By the proposed cluster



Figure 22

nomenclature this compound is named: 4,5,4-tridecabenzocluster. (The IUPAC name is: phenanthro(3,4,5,6-bcdef)ovalene.) Similarly, the compound shown in Figure 23 (common name - circobiphenyl) would be named 2,3,2,3,2-dodecabenzocluster, rather than the IUPAC name: naphth(2',1',8',7':4,10,5)-anthra-(1,9,8-abcd)-coronene.

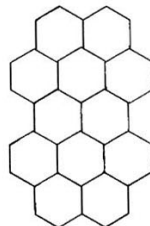


Figure 23

Consider the compound $C_{40}H_{20}$ shown in Figure 24. This polycyclic is formed from only ten benzene rings without the connecting bonds between IUPAC positions 17-20 and 18-19;

thereby having four additional hydrogen atoms. This compound, which may be called: 4,1-3x-1,4--decabenzocluster, by IUPAC nomenclature would be: 15,21:16,20-dimethenaphtho-(1,2-a)-phenanthro-(3,4~~l~~)-naphthacene. Using the circum prefix it would not be possible to name this compound since there exists no common name for the skeletal structure without the connecting bonds. The linear nomenclature described in Part I is: jhgfdbaa-decabenzene.

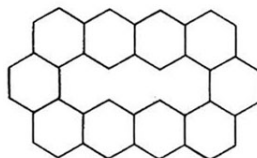


Figure 24

One additional polycyclic aromatic hydrocarbon that requires consideration is graphite -- the only hydrocarbon containing NO hydrogen. As the plane is covered by fused benzene rings (refer to the inner rings of Figure 22) notice that none of the internal carbons (i.e., none of the double or triple points) contain a bonded hydrogen atom. In other words, the hydrogens are pushed toward the outer boundary -- which is constantly receding as more rings are added. Since the geometric picture of graphite is that of an infinite planar tessellation, the hydrogens "at infinity" are, for practical purposes, non-existent. Thus, graphite may be viewed as an infinite (in two dimensions) benzene cluster.

III. SUPPLEMENTARY INFORMATION

6. CHARACTERISTIC NUMBER

A unification of the different possible hexagonal configurations may be introduced at this point: Consider the sum of each of the uninterrupted rows of hexagons (horizontal, diagonally left and diagonally right). Count interrupted rows as separate entities. For Figure 25 (eeba-hexabenzene) horizontally there exist rows of 3, 2 and 1 ring; diagonally left there exists one row of three rings and three rows of one ring; diagonally right there exist two rows of two rings and two uninterrupted rows of one ring (actually one interrupted row). By subtracting one from each of these entries, the "characteristic number" X is formed:

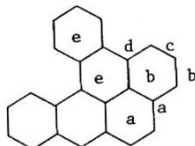


Figure 25

$$X = \frac{(3-1)+(2-1)+(1-1)}{\text{Horizontal}} + \frac{(3-1)+3(1-1)}{\text{Diag. Left}} + \frac{2(2-1)+2(1-1)}{\text{Diag. Right}} = 7$$

This number plus one equals the number of rings (R) plus the number of triple points (T) for all configurations studied, with the possible exception of coronene. In the case of coronene, and all configurations in which an additional ring has been enclosed, there are two satisfactory adaptations:

- (1) If coronene is to be considered a six ring compound, we may adjust the characteristic number by adding the concept of genus (G) from elementary topology:

$$X = R + T + G - 1$$

$$R = 6, T = 0, G = 1$$

$$X = 2(2-1)+2(1-1)+2(2-1)+2(1-1)+2(2-1)+2(1-1)=6$$

- (2) If coronene is to be considered a seven ring compound:

$$R = 7, T = 6, G = 0$$

$$X = (3-1)+2(2-1)+(3-1)+2(2-1)+(3-1)+2(2-1) = 12$$

Table III illustrates the computation of the characteristic number of the various aromatic hydrocarbons through the tetrabenzene, as well as some more complicated configurations such as portrayed above in Figures 21 through 24.

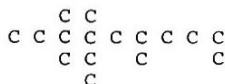
TABLE III
CHARACTERISTIC NUMBERS

Synthetic or Cluster Name	Reference	Horizontal Rings - 1	Diagonal Left - 1	Diagonal Right - 1	X = R + T + G - 1
Benzene		1 - 1	1 - 1	1 - 1	0 = 1 + 0 + 0 - 1
Dibenzene	Figure 5	2 - 1	2(1 - 1)	2(1 - 1)	1 = 1 + 0 + 0 - 1
Tribenzene	Figure 8				
A-		3 - 1	3(1 - 1)	3(1 - 1)	2 = 3 + 0 + 0 - 1
B-		2 - 1	3(1 - 1)	(2-1)+(1-1)	2 = 3 + 0 + 0 - 1
C-		2 - 1	2 - 1	2 - 1	3 = 3 + 1 + 0 - 1
Tetrabenzene	Table I				
AA-		4 - 1	4(1 - 1)	4(1 - 1)	3 = 4 + 0 + 0 - 1
BA-		(3-1)+(1-1)	4(1 - 1)	(2-1)+2(1-1)	3 = 4 + 0 + 0 - 1
CA-		(3-1)+(1-1)	(2-1)+2(1-1)	(2-1)+2(1-1)	4 = 4 + 1 + 0 - 1
BB-		(2-1)+2(1-1)	(2-1)+2(1-1)	(2-1)+2(1-1)	3 = 4 + 0 + 0 - 1
CB-		2(1 - 1)	(2-1)+2(1-1)	2(1 - 1)	5 = 4 + 2 + 0 - 1
DB-		2(2 - 1)	4(1 - 1)	(2-1)+2(1-1)	3 = 4 + 0 + 0 - 1
FB-		(2-1)+2(1-1)	(2-1)+2(1-1)	(2-1)+2(1-1)	3 = 4 + 0 + 0 - 1
2x-3, x-2x, 3-4x Nonabenzocluster	Figure 21	3(3-1)	2(2 - 1)	(3-1)+2(2-1)	12 = 9 + 4 + 0 - 1
4, 5, 4-Tridecabenzocluster	Figure 22	(5-1)+2(4-1)	3(3-1)+2(2-1)	3(3-1)+2(2-1)	26 = 13 + 14 + 0 - 1
2, 3, 2, 3, 2-Dodecabenzocluster	Figure 23	2(3-1)+3(2-1)	2(4-1)+2(2-1)	2(4-1)+2(2-1)	23 = 12 + 12 + 0 - 1
4, 1-3x-1, 4-Decabenzocluster	Figure 24	2(4-1)+2(1-1)	2(2-1)+6(1-1)	2(2-1)+6(1-1)	10 = 10 + 0 + 1 - 1

7. THE SQUARE TESSELLATION AND ALKANES

Consider a sequence of some letter (say C) in the center of squares of a piece of graph paper

(Figure 26). If these squares are "rook-wise" contiguous -- every marked square can be joined to every other



marked square by a horizontal and/or

Figure 26

vertical sequence of straight line segments passing only thru marked squares, this may be viewed as the skeletal structure of an alkane PROVIDED there are no closed loops containing unmarked

squares. Using IUPAC nomenclature¹, Figure 26 represents a decane (despite that only 9 C's appear in a single row. The C at the bottom right is rook-wise contiguous to the long row.)

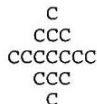
The name of this compound is: 4-Ethyl-3,3,4,6-tetramethyldecane. Although the alkanes are three dimensional molecules, the two-dimensional picture is nearly adequate since each carbon atom has four ligands and each square has four rook-wise contiguous squares. In practice, however, the presence of tertiary butyl or other highly branched sub-groups can create

confusion for the two dimensional picture since a common edge of the tessellated squares implies a chemical bond. Consider 4,4-Di-tert-butylheptane shown in Figure 27. If the connecting bonds had been omitted from the picture and the

positioning not distorted to falsely imply that there was double the distance between certain

carbon atoms and not between others, the picture would look like Figure 28. Notice that Figure 28

is named: 4,4-Diethyl-3,3,5,5-tetramethylheptane -- a drastically different compound. In other words, the



square tessellation is a rather imperfect model, due mainly to the mapping of a three dimensional model onto a rigid two dimensional structure.

Figure 28

8. THE TRIANGULAR TESSELLATION AND HEXA-COORDINATED COMPOUNDS

When considering the triangular tessellation note that the problems associated with square tessellations as a model for alkanes is further exacerbated by the need to introduce perspective for six ligands. In

Figure 29, the central sulfur atom has six ligands -- indicated by their location on a cubic coordinate system mapped perspectively. In turn, each of these ligands can have from one (connecting back to the sulfur atom) up to six ligands about it. However, the ligand lying on the positive x-axis (denoted by L_x) is not normally

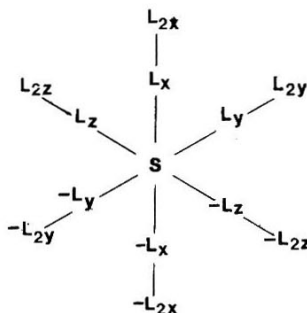


Figure 29

connected to the ligand lying on the positive y-axis (denoted L_y), despite the fact that the distances seem to imply this. In order to prevent this interpretation a centro-perspective scheme in which the distances to the center are doubled (such as in Figure 30) or some equally

exotic picture would be required. Thus, despite the fact that the cubic tessellation might be the desired geometry, the disadvantages of a planar pictorial representation outweigh the utility of the tessellation approach.

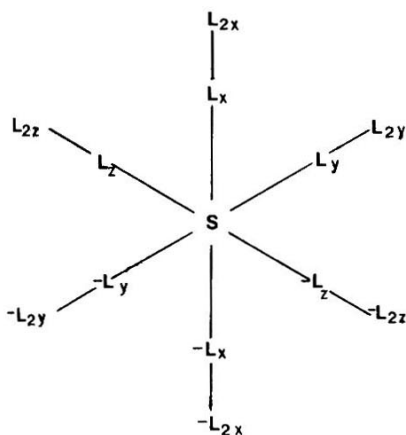


Figure 30

APPENDIX

ALGORITHM # 1

(used in synthesis process)

1. Start with the longest chain. Make this chain horizontal. If there exist two or more equally long chains, see below:
2. Position unsymmetric forms in the first quadrant so that they are as nearly centered as possible, but with any excess being top heavy to the upper right. Some criterion to determine comparative "top heaviness", such as the location of the center of gravity, must be employed for larger molecules.
3. Position A elongates the longest chain. When position A could be either an end chain or an internal chain, choose the free end of the end chain. If there are two free ends, choose the one to the left. (See #2 above).
4. Position B is the first edge clockwise to A, etc.
5. Skip a position in lettering when a configuration is obviously redundant.
6. For each ring being added, repeat Steps 1 through 5.

APPENDIX

ALGORITHM # 2

(used in analysis process)

1. Consider all possible rigidized forms of the molecule. Analyze the various configurations as follows:
2. Starting from the longest chain, the first two rings of this chain are considered part of the skeletal structure and are not named. Designate these two rings with an asterisk. All other rings in this chain are named a.
3. In addition to the rings in the longest chain, also designate as position a the extensions (in both directions) of this longest chain. For n equal length longest chains, there exist 2n possible locations for a.
4. Choose as b the first edge either counterclockwise or clockwise to a. Thus there exist 4n possible locations for b. Continue "walking" around the molecule naming edges c, d, etc.
5. Of all the "interim" names described above, choose the lexicographically lowest -- reading right to left.

LITERATURE CITED

1. International Union of Pure and Applied Chemistry, "Nomenclature of Organic Chemistry," Sections A, B, C, D, E, F, G, H. Revised Edition, Pergamon Press, Oxford, England, 1977.
2. Elk, Seymour B., Carib. J. Science & Math, 1968, 1, 1-30.
3. Clar, Eric, "Polycyclic Hydrocarbons, I & II," Academic Press, New York, 1964.
4. Harary, Frank, "Graph Theory," Addison-Wesley, Reading, Mass., 1971, p. 64-5.
5. Sloane, Neil J. A., "A Handbook of Integer Sequences," Academic Press, New York, 1973, p. 102 (Reference # 1072).
6. Klarner, David A., Can J. Math, 1967, 19, 851-863.
7. Clar, Eric, "The Aromatic Sextet," John Wiley & Sons, London, 1972.
8. Courant, Richard & Robbins, Herbert, "What Is Mathematics?" Oxford University Press, London, 1961, 256-259.