

## A CANONICAL ORDERING OF POLYBENZENES AND POLYMANTANES USING A PRIME NUMBER FACTORIZATION TECHNIQUE

Seymour B. ELK

*Elk Technical Associates, 321 Harris Place, New Milford, NJ 07646, USA*

### Abstract

The correlation described over twenty years ago by Matula between the prime factorization of integers and the class of alkanes is re-examined with a view to explaining the probable reason why there have, to date, been no major extensions of this idea. By considering the class of alkanes as a one-dimensional one-parameter system, a new perspective on the method is gained that is amenable to extension, but in a different direction than originally anticipated. With this new perspective, the classes of polybenzenes and polymantanes are seen to be the representatives of two- and three-dimensional one-parameter systems, respectively. A "nomenclature", comparable to one that Matula used for alkanes, is created that gives a unique canonical name to all possible combinations of either polybenzene or polymantane modules. Such a "nomenclature" contains a "built-in" means of positioning the molecule in the field of interest in accordance with arbitrary pre-selected criteria, such as Patterson's rules, and also coding that indicates symmetries inherent in the structure of this "molecule".

### 1. Introduction

A little over twenty years ago, Matula [1] formulated a one-to-one and onto correlation between the prime factorization of numbers and the class of rooted trees, and then noted that the class of rooted trees had a nearly one-to-one relation with the alkanes. (Nearly, because any integer whose prime factorization is the product of five or more primes is not a viable alkane; i.e. 32 or 48, etc. would require five bonds at a carbon atom.) Nevertheless, despite the fact that these relations could be used in formulating an ordering and, thus, a canonical "nomenclature" for alkanes, so far as we know, all attempts to meaningfully expand this system have been fruitless. This is not to say that no extensions have been made. One direct extension of the system which we have played with previously, but at the time did not deem significant enough to follow through, is created by dividing the prime number set into three sets:

$$S_1 = \{2, 7, 17, \dots\}, S_2 = \{3, 11, 19, \dots\}, S_3 = \{5, 13, 23, \dots\}.$$

This allows for the creation of a numbering and naming of all acyclic hydrocarbons by using elements of  $S_1$  for single bonds,  $S_2$  for double bonds and  $S_3$  for triple bonds – even though it does produce much larger numbers. Figure 1(a) shows a dodecane and

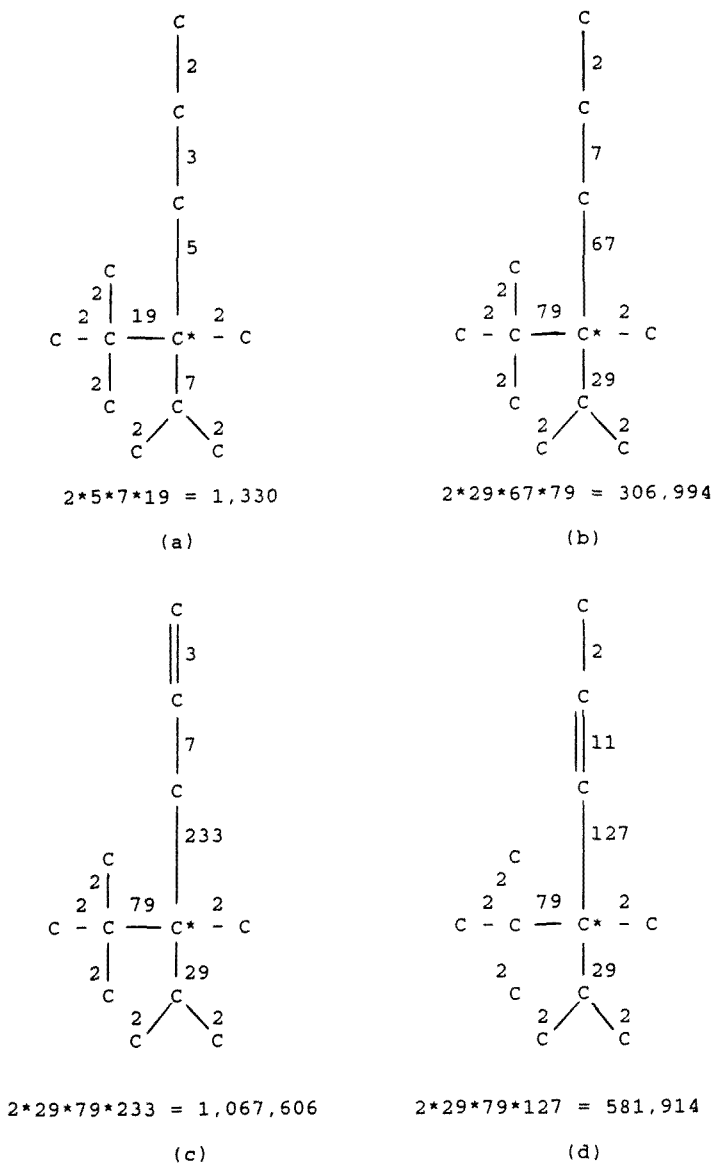


Fig. 1. Canonical labelling of a selected dodecane and two related dodecanes.

Matula's labelling – which gives the canonical name 1330. Figure 1(b) shows that the same dodecane is named 306,994 when the method is modified to allow for multiple bonds. Similarly, in the expanded version, the location of a multiple bond is shown: in figs. 1(c) and 1(d), a different one of the single bonds has been replaced by a double bond, producing the names 1,067,606 (c) and 581,914 (d), respectively.

In order for this method to be useful, it had generally been assumed that next we must treat cyclic compounds, i.e. some means of reducing the set of cyclic compounds to that of acyclic subsets by means of "cut-points", etc. must be found. We propose a different direction of attack.

An examination of the class of alkanes shows it to be a one-parameter binary system; i.e. once a single binary set (in this case a Hadamard incidence matrix of the carbon atoms) has been given, all other specifications are known. Other one-parameter systems abound in chemistry; thus, instead of modifying Matula's method to apply to different bond order in acyclic hydrocarbons, we shall apply the technique to different parameters. In particular, we shall focus on a prototype one-parameter system in each applicable dimension. Just as the acyclic alkanes were the one-dimensional prototype, polybenzenes shall be the two-dimensional and polymantanes the three-dimensional prototypes, respectively.

Although application to any parameter using a modification patterned after the above expansion to all acyclic hydrocarbons will produce large numbers, the size of the numbers produced will be a function of the graph-theoretical radius from the designated node. For single atoms, as in the acyclic hydrocarbons, very large radii molecules are of practical importance; however, if a different module is the basis of description, the size of the moieties of pragmatic (and also theoretical) interest may well be within practical limits.

## 2. Polybenzenes

In order to formulate a canonical ordering of polybenzenes, let us focus on the hexagonal tessellation of the plane. First of all, we note that each hexagon is surrounded by six other hexagons. There are now two different approaches to consider: based on either the local or the global geometry of the molecule under consideration. These roughly correspond to the synthetic and analytic nomenclature discussed in ref. [2]. In both instances, we shall select one ring, named the "root", and designate each of the other hexagons of the molecule using two parameters: the graph-theoretical distance (abbreviated GTD) from the root and an indicator of orientation. The indicator of orientation shall be either purely an ordered polar coordinate system using distance first and then, for equal distance, angle measure, or, more conveniently, angular preference is pre-selected by some arbitrary criterion, such as Patterson's rules [3]. In this latter case, the location of each hexagon being annellated may be indicated by one of six subsets as follows:  $\{S_1\}$  corresponds to annellation on a horizontal line to the right of the root,  $\{S_2\}$  to annellation in the "right up" position,  $\{S_3\}$  to "right down",  $\{S_4\}$  to "left up",  $\{S_5\}$  to "left down" and  $\{S_6\}$  to "left horizontal". This is shown in fig. 2.

Note that in the intuitively simpler synthetic nomenclature method, we started with a single hexagon and annellated hexagons one at a time to synthesize the desired molecule, while in the analytic nomenclature method, we considered the orientation of

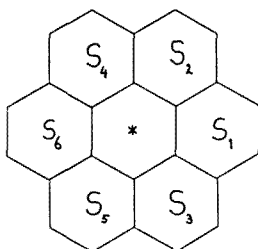


Fig. 2. Labelling hexagons to conform to the intent of Patterson's rules.

the molecule as immutable and then examined the set of all possible orientations before choosing some "optimal" one. We shall use comparable procedures here.

### 2.1. THE "SYNTHESIS" METHOD

For the "synthesis" method, we shall be focusing on just on ring at a time. For the first step, we focus on the "root" (indicated by a star). To this root, we may annellate from one to six additional rings depending on how many of these rings are in the target molecule. We next focus on each of the rings that we have annellated to the root and consider the orientation of these rings to the root. We then proceed outward (increasing GTD), using each ring added as the new center, until the entire molecule has been marked.

To illustrate how this step of the method works, let us now apply this technique to a representative polybenzene: figure 3 is the five-ring system with IUPAC name dibenzo[c, 1]phenanthrene – which was named FBB-pentabenzene in the synthetic

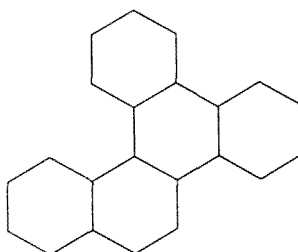


Fig. 3. A five-ring system.

method of ref. [2] and (coincidentally) fbb-pentabenzene in the analytic scheme, and, using the method of hexagonal tessellations in ref. [4], 0163-pentabenzoccluster. Note that all of these names (IUPACs and our three different ones) contain a sequence of descriptors, instead of a single unique descriptor, which can be used for sequentially ordering the entire set of polybenzenes. To create such a single descriptor, henceforth called a "name", select any single ring as the reference and then number the rings

according to the GTD from the referenced ring. This is shown in fig. 4 for the target molecule in the orientation that it was given. The individual rings are further named with a subscript indicating which of the six subsets is relevant.

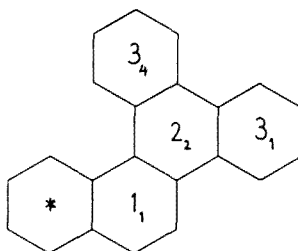


Fig. 4. One possible labelling of the sample ring system.

As the next step, let us sub-divide the set of prime numbers into six subsets corresponding to sets  $S_1$  through  $S_6$  (see table 1). Multiplication of the numbers in table 1, which correspond to each hexagon in the target molecule, gives a unique

Table 1

Allocation of prime numbers with respect to distance and orientation;  
synthesis method for polybenzenes

Distance	$\{S_1\}$	$\{S_2\}$	$\{S_3\}$	$\{S_4\}$	$\{S_5\}$	$\{S_6\}$
1	2	3	5	7	11	13
2	17	19	23	29	31	37
3	41	43	47	53	59	61
4	67	71	73	79	83	89
5	97	101	103	107	109	113
6	127	131	137	139	149	151

reversible name:  $2 * 19 * 41 * 53 = 82,574$ . Note that any one of the five rings could have been chosen as the reference and any of the twelve possible orientations could have been chosen. Consequently, we desire to choose the smallest number from this set of sixty possibilities.

Although the above method may seem like a formidable search problem and one that will grow even more formidable with an increasing number of rings, it is greatly shortened by noting that a different name is generated for each choice of root in the same molecule. Consequently, using the principle that the optimal name is the one which has the smallest number, we see that the only reasonable rings to choose as the starting hexagon are those at the center of the system, since these will have the lowest numbers to be multiplied. (In the preceding sentence, the word "center" is defined graph-theoretically [5].) Thus, there are only two rings which could qualify for center in this example. (In general, the maximum number of rings which could be the "center"

is three – a number that can be realized only in selected peri-condensed systems.) In a similar manner, the choice of which of these at most three rings to choose as the center is the one having the minimum sum of the distances to the other rings. This is seen to be six in fig. 5(a) versus only five in fig. 5(b). The concomitant names are:  $3 * 13 * 17 * 29 = 19,227$  and  $2 * 7 * 11 * 37 = 5698$ , respectively. Additionally, the twelve orientations

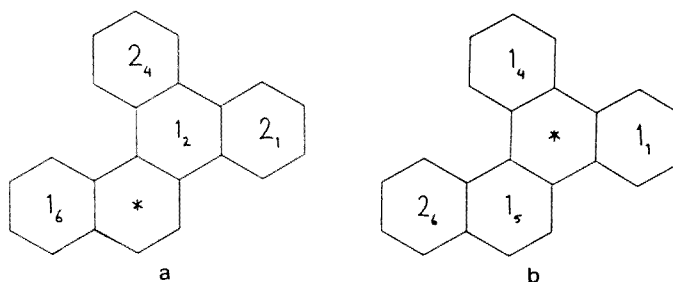


Fig. 5. Labelling of the sample ring system starting from a "center" ring.

may be selectively reduced by maximizing the number of rings from  $\{S_1\}$  and minimizing the number from  $\{S_6\}$ . Note that a lower number is produced by reflecting the molecule shown in fig. 5(a) in a plane and rotating it  $120^\circ$  (fig. 6). This produces the name:  $2 * 7 * 11 * 19 = 2926$ , which is the minimum of the twenty-four possible names that could be produced and shall be used as the canonical name in accordance with the algorithm given in table 2.

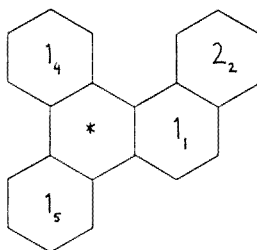


Fig. 6. Canonical labelling of the sample ring system – after reflection and rotation.

One major advantage of this technique is its universality. It is not limited to cata-condensed [6] ring systems, or even to chemically viable ring combinations. In fig. 7, for example, which has an odd number of triple points [7] and is thus chemically unrealizable, this is still an easy molecule to name. First of all, note that two choices of center and of orientations have the same minimum sum: 8. Using the orientation given:  $2 * 7 * 11 * 13 * 17 * 19 = 646,646$  (fig. 7(a)) and in the orientation that yields a canonical name:  $2 * 3 * 5 * 13 * 29 * 37 = 418,470$  (fig. 7(b)). Furthermore, we may change our focus and use this technique to orient a molecule in such a way that it

Table 2

## Algorithm to generate the synthesis names for polybenzenes

- 
- Step 1.* Locate the center ring(s) of the system.
- Step 2.* In each hexagon of the system, mark the graph-theoretical distance to the center ring. Repeat this step once or twice if there are two or three center rings.
- Step 3.* Establish an orientation between rings of distance  $n + 1$  with respect to all contiguous rings of distance  $n$ , and number each such ring with a subscript to indicate its orientation with respect to the lower distance contiguous ring. The subscripts are: 1 = horizontal-right; 2 = up-right, 3 = down-right; 4 = up-left; 5 = down-left; 6 = horizontal-left.
- Step 4.* In the event that there are two or more paths to reach a given ring, consider only that subset having the minimum graph-theoretical distance. Select the one whose first difference in ring designation has the lower orientation.
- Step 5.* Assign to each ring of the system the prime number given in table 1.
- Step 6.* The product of the numbers inscribed in the rings in step 5 is the desired name. If there was more than one center ring, the canonical name is the minimum of names generated.
- 

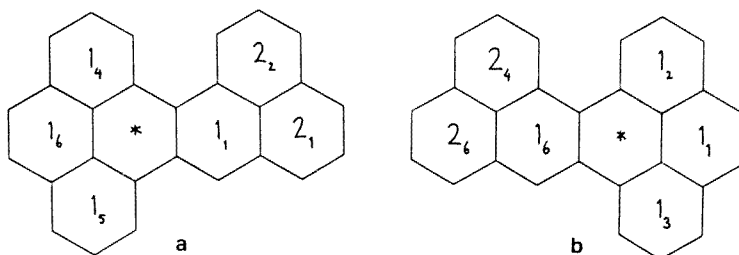


Fig. 7. Sample labelling of a chemically unrealizable ring system.

conforms to whatever intuitive concept of "centering", such as based on the center of gravity of a flat plane of uniform hexagonal tiles, etc., that we desire. In other words, with respect to the heuristic intent that we feel underlies IUPAC's rules of nomenclature [8], we have another tool to use. This is yet another perspective on earlier examinations that we have made on this subject [9].

In the event that more than one path of a given length can be used with a given hexagon, all possibilities must be examined. This, however, is not a problem since, because of the density of the prime numbers, the lowest first number will yield the lower number for the final path. This is illustrated in fig. 8, where the upper leftmost hexagon has two paths – through  $1_4$  and  $1_6$ . The first of these names this hexagon  $2_6$ , while the second names it  $2_4$ . The resultant name using  $1_4$  followed by  $2_6 = 7 * 37 = 259$ , which is less than  $1_6$  followed by  $2_4 = 13 * 29 = 377$ . (Note that we did not need to examine the entire molecule.)

Problems with the synthesis method arise, however, since there is not the desired 1:1 relationship. This is seen in fig. 9, since the ring marked with just a distance 2 and no subscript would be named  $2_2$  if  $1_1$  is the reference, but would be  $2_1$  if  $1_2$  is the reference. In other words, a different number would be used in the multiplication

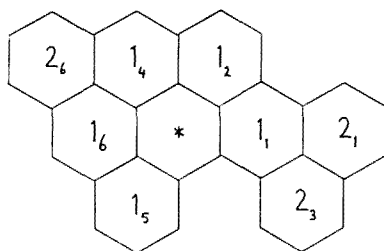


Fig. 8. When alternative paths create a potentially ambiguous combination of hexagons.

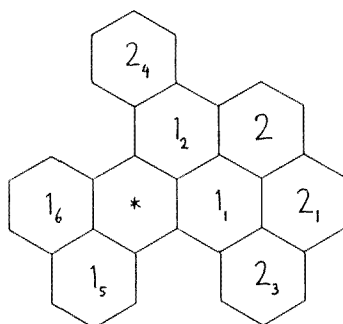


Fig. 9. A combination of hexagons that creates further ambiguity.

process to determine the canonical name. Furthermore, in the factorization process that regenerates the individual rings, we would not know which ring was being specified by  $2_2$  without a sequence rule that could not maintain the desired consistency. This is an analogous situation to the one that led to the need for the analytic method in ref. [2].

Before examining the analysis method, note that we may start from the set of integers and generate all possible polybenzenes – in much the same way as we could generate Matula's correlation to alkanes. Similarly, this relation is not one-to-one; the same structure may be generated for several numbers. This is illustrated in table 3 for numbers 1 through 50. Furthermore, we observe that no prime raised to higher than the first power was found to be useful at this time.

Another contradiction in the synthesis method is seen by considering the integer 51. Here, the set of hexagons that 51 refers to is disconnected and, thus, for pragmatic purposes shall not be treated by this technique. Attempts to apply this method to ring assemblies create contradictions. Although fig. 10(a) seems to allow 51 to be represented by a naphthalene–benzene ring assembly, we should have to disallow reorientation of any molecule having a ring assembly. This is true, since such a ring assembly could now be reoriented as in fig. 10(b). This would imply that the benzene ring of the ring assembly should be designated as  $2_2$ , and that the three rings of the system were the root,  $1_1$  and  $2_2$ ; i.e. the molecule would be named 38. However, the name 38 would also refer to B-tribenzene with the root in the leftmost ring, as in fig. 10(c). Since this



Table 3  
 Tabulation of synthesis name of polybenzenes for integers from 1 to 50

Number	Synthetic name [2]
1	Benzene
2 = 3 = 5 = 7 = 11 = 13	Dibenzene (prime with GTD = 1)
4, 8, 9, 12, 16, 18, 20, 24, 25, 27, 28, 32, 36, 40, 44, 45, 48, 49, 50	Not used; involves a higher power of a prime
6 = 10 = 15 = 21	C-tribenzene
14 = 22 = 39 = 46	B-tribenzene
17, 19, 23, 29, 31, 37, 41, 43, 47	Not used; prime with GTD > 1
26 = 33 = 35	A-tribenzene
30 = 42	CB-tetrabenzene
34	Compare with 26; same structure, different root
38	Compare with 14; same structure, different root

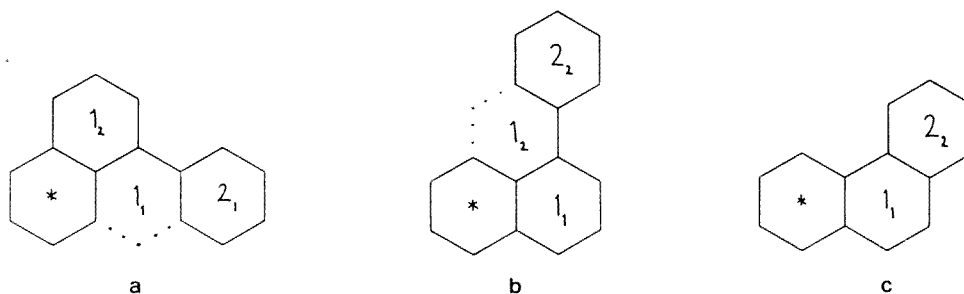


Fig. 10. Contradictions encountered with the synthesis method (see text for details).

is untenable, we disallow naming ring assemblies using the "synthesis" method. Although the same problem should have arisen had we closely scrutinized the name 38, the presence of an adjacent ring to annellate obscured our noting this problem at that time.

## 2.2. THE "ANALYSIS" METHOD

Let us now turn to the method that is modelled after the analytic method of ref. [2]. This method selects one henceforth immutable orientation for a given polybenzene and then designates distance and orientation based on this reference. In this method, the reference hexagon is surrounded by equidistant rings of hexagons. In its simplest form, there are  $6d$  hexagons in each ring, with  $GTD = d$ . A grid of such hexagons is shown in fig. 11. Table 4 lists the distance versus prime number relationship, while table 5 lists the analysis algorithm. These are the analysis method analogs of tables 1 and 2, respectively. The analog to table 3 will not be given yet, since in its

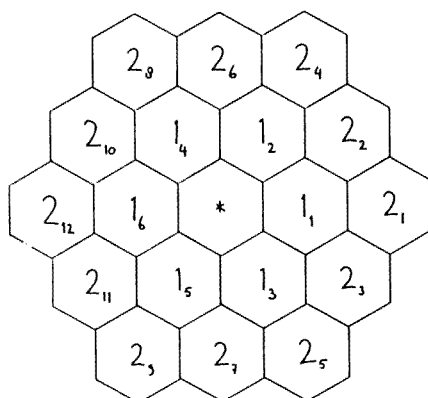


Fig. 11. A grid of equidistant hexagons.

Table 4

Allocation of prime numbers with respect to distance and orientation;  
analysis method for polybenzenes

Distance						
1	2	3	5	7	11	13
2	17 19	23 29	31 37	41 43	47 53	59 61
3	67 71 73	79 83 89	97 101 103	107 109 113	127 131 137	139 149 151

Table 5

Algorithm to generate the analysis names for polybenzenes

- Step 1.* Locate the root(s) of the system.
- Step 2.* Starting from the root, mark the GTD distance to every other hexagon in the system. Repeat this step once or twice if there are two or three roots.
- Step 3.* Orient the molecule in any of the 12 possible positions, or if there is  $n$ -fold rotational symmetry, the  $12/n$  positions.
- Step 4.* Number each of the marked hexagons with a subscript to indicate its orientation with respect to the center. For molecules lacking rotational symmetry, the subscripts in each of the circumscribing ring of hexagons are:  
1 = horizontal-right; even numbers ( $2n$ ) correspond to the  $n$ th hexagon measured counterclockwise from hexagon 1; odd numbers ( $2n + 1$ ) correspond to the  $n$ th hexagon measured clockwise from hexagon 1, up to  $6d$ , where  $d$  is the graph-theoretical distance from the center hexagon.  
For molecules possessing rotational symmetry, the subscripts are numbered counterclockwise from 1 = horizontal-right.
- Step 5.* Assign to each ring the appropriate sequential prime number. For the non-rotationally symmetric case, this is given in table 4.
- Step 6.* The product of the numbers inscribed in the rings in step 5 is the desired name for the specific center and orientation.
- Step 7.* Repeat steps 4 and 5 for each possible orientation and center; the canonical name is the minimum of names generated in step 6.

present form this method, while fulfilling the consistency requirement it was designed to treat, is very inefficient for molecules possessing rotational symmetry.

Instead of having to discard all higher powers of a prime number, it is efficient to use primes raised to the second power for  $C_2$  symmetry, to the third power for  $C_3$  symmetry and to the sixth power for  $C_6$  symmetry. Two items to note at this point are:

- (1) These are the only possible types of rotational symmetry that are applicable for polybenzenes. In other words, we would have no polybenzenes that are represented by a prime number raised to the fourth or fifth power, or to the seventh or higher power.
- (2) Even though it would be highly desirable, we have not taken into consideration reflectional symmetry. In other words, we have eschewed the assignment of the unused fourth or fifth powers of the primes for this purpose, feeling that such a usage would be counterproductive. Instead, we prefer to maintain the relationship that the number of rings in a given molecule can always be determined from the name – it is one more than the number of primes that were used in the name.

With the above usage of primes raised to powers, it is a simple matter to rewrite table 4 to indicate each of the rotational symmetries, and to amend table 5 to first determine the symmetry and then use the appropriate prime numbers to determine the desired name. Furthermore, having done this, we may now determine such symmetry by noting whether the name given is a perfect square, a perfect cube or a perfect sixth power. This is illustrated in fig. 12 for the 13-ring system possessing sixfold symmetry.

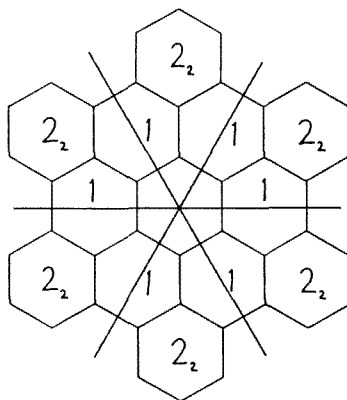


Fig. 12. A 13-ring system with sixfold symmetry.

Note that at  $GTD = 1$ , all six hexagons are equivalent and so only the first prime must be reserved for this distance. Next, for  $GTD = 2$ , the prime number 3 corresponds to linear annellation – which is not present in the target molecule, while prime number 5 corresponds to angular annellation – which is. Thus, all 13 rings are now covered by: 6 rings marked 1, 6 marked  $2_2$  and the starred ring. Consequently, the desired name is  $2^6 * 5^6 = 1,000,000$ . In a similar manner, if we had noted only the

threefold symmetry of this molecule, it would be given the non-canonical name  $(2 * 3 * 7 * 13)^3 = 162,771,336$ . This is equivalent to saying that the canonical name is coded to include all rotational symmetries, but that there may exist other valid (but not canonical) names which describe only some or none of the symmetries. Using symmetries, even esoteric combinations, such as those used by Brunvoll, Cyvin and Cyvin [10], are easy to code. Table 6 is the analysis method analog to table 3 and gives a translation of each of the smaller integers into its simplified analysis name.

Table 6  
Tabulation of analysis name of polybenzenes for integers from 1 to 50

Number	Analytic name [2]
1	Benzene
2 = 3 = 5 = 7 = 11 = 13	Dibenzene (prime with GTD = 1)
4 = 9 = 25 = 26 = 33 = 34 = 35	a-tribenzene
6 = 10 = 21	c-tribenzene
8 = 27	jb-tetrabenzene
12 = 18 = 20 = 28 = 44 = 45 = 50	Square factor, but not perfect square
14 = 15 = 22 = 38 = 39 = 46	b-tribenzene
16, 32, 48	Uses 4th or 5th power of a prime
17, 19, 23, 29, 31, 37, 41, 43, 47	Prime number with GTD > 1
30 = 42	eb-tetrabenzene
24 = 36	ica-pentabenzene
40	Cube factor, but not perfect cube
49	Too large a perfect square

### 3. Polymantanes

We now turn our attention to polymantanes. Of the four possible methods of fusing two adamantane modules (fig. 13), only the face-fusion of modules (fig. 13(d)) is capable of tessellating the three-dimensional embedding space that this class of molecules lie in; thus, we shall limit our attention to such fusions. Consequently, we can develop a somewhat simpler classification scheme than Balaban and Schleyer did in ref. [11]: Each face-fusion of two adamantane modules is an equilateral skew hexagon such that the line segments between opposite vertices are concurrent in a point that is the center of the circumsphere of that skew hexagon. We may, therefore, completely describe the polymantane by studying the set of circumcenters. Note that adamantane is represented by the empty set, diamantane by a single point and triamantane by a line segment connecting the two circumcenters. Since each of these figures (empty set, point and line segment) is unique, there is only one isomer each for adamantane, diamantane and triamantane. Anellating a fourth adamantane module to triamantane can be accomplished at three distinct locations, producing a circumcenter set that is collinear

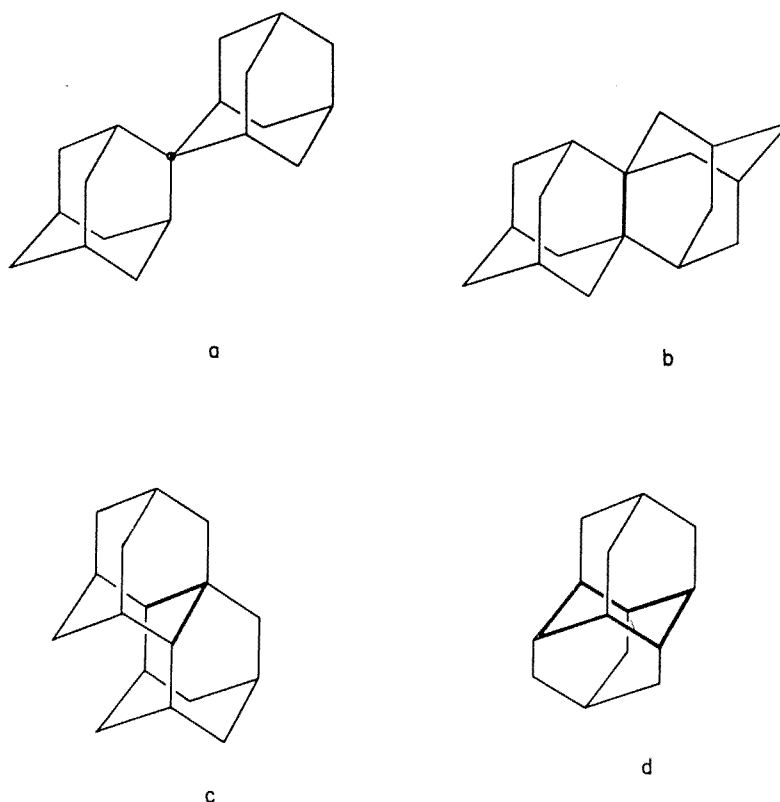


Fig. 13. Four methods of fusing two adamantane molecules.

(A-tetramantane), angular (B-tetramantane) and triangular (C-tetramantane) – a synthesis process that is equivalent to the one used for the hexagonal modules in a polybenzene [12]. Thus, in an identical way, we may find the center of the circumsphere and postulate an order for the four planar-fused faces. In other words, the set of prime numbers would be divided into four subsets and would generate tables which are analogous to those used for the polybenzenes. Because of the time and space constraints of both the symposium presentation and this report, and because polymantanes of the size required to encounter the problems that we did for polybenzenes are probably still a long way in the future, the corresponding details required for formal algorithms have not as yet been examined.

## References

- [1] D.W. Matula, *SIAM Rev.* 10(1968)273.
- [2] S.B. Elk, *MATCH* 8(1980)121–159.
- [3] A.M. Patterson, *J. Amer. Chem. Soc.* 47(1925)543–561.
- [4] S.B. Elk, *MATCH* 17(1985)255–268.
- [5] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969), p. 35.

- [6] A.T. Balaban and F. Harary, *Tetrahedron* 24(1968)2506.
- [7] *Ibid.* 2(1980)126.
- [8] *IUPAC Nomenclature of Organic Chemistry*, section A (Pergamon Press, Oxford, 1979).
- [9] S.B. Elk, *Polycyclic Aromatic Compounds* 1(1990)109–121.
- [10] J. Brunvoll, B.N. Cyvin and S.J. Cyvin, *J. Chem. Inf. Comput. Sci.* 27(1987)171–177.
- [11] A.T. Balaban and P.vR. Schleyer, *Tetrahedron* 34(1978)3599–3609.
- [12] *Ibid.* 2(1980)125–127.