

A Canonical Assignment of Locant Numbers to Fisular Compounds—Especially Fullerenes—Based on Graph Theoretical Principles

Seymour B. Elk

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

Received October 5, 1994[®]

A canonical assignment of locant numbers to fisular compounds—especially fullerenes, is proposed using principles common to graph theory. The foundations of the assignment scheme is that of first minimizing the graph theoretical distances (GTDs) from a selected atom of the fisular cage and then minimizing GTDs from the next lowest number atom, etc.

1. INTRODUCTION

The method of assigning canonical names to organic molecules that has evolved over the years began with the initial expectation that nearly all important molecules would be linear chains of connected carbon atoms. The system first codified by the Geneva Convention in 1892 has worked moderately well for most compounds which we now categorize as “acyclic”—even though we note that the actual bond angles for the VAST majority of these compounds are much closer to the tetrahedral angle ($\sim 109^\circ 28'$) or the trigonal angle (120°) than to linearity (180°). The nomenclature of such chains was now to simply label one end of the chain as locant no. 1, and each successive carbon in the chain with the corresponding integer. In other words, the nomenclature algorithm required only a means to decide which, if there were more than one, chain to choose and which end of the (selected) chain to start with. This is standardized in Rule A-2 in the International Union of Pure and Applied Chemistry's (IUPAC) “Nomenclature of Organic Chemistry”.¹

Before progressing further, we note that near unanimity of opinion as to which chain to pick and the optimal locant numbering of that chain is achieved ONLY for (1) all-carbon-atom, (2) nonbranched, (3) fully-saturated, (4) uncharged, (5) etc., chains. In other words, a truly one-parameter system is required. This is illustrated in Figure 1, wherein the double bond takes precedence in I.U.P.A.C.'s¹ scheme (2-ethylbutene) vs the longest chain having primacy in other systems. For example, in nodal nomenclature,² this molecule would be named 3-methenylpentane, etc. The next major advance in our perception of molecules was the recognition that some chains might be closed to form cycles. The presence of such “rings”³ now introduces more parameters that must be standardized if we are to have a canonical system as well as the need for, at least, a two-dimensional embedding space. (This second requirement can be partially finessed by the introduction of standardized symbols for commonly occurring modules, such as has been done in the Wiswesser notation;⁴ however, a high price is exacted in that deviations from the selected standards usually cannot be made without undesirable complications.)

The first, and most obvious, choice of parameter in a ring is to select the starting atom. Note that any of the atoms of the ring, rather than just the two ends of a chain, may be

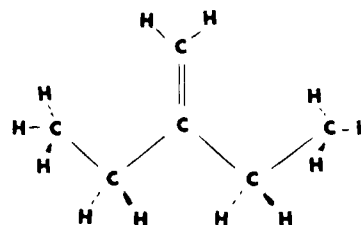


Figure 1. 2-Ethylbutene (I.U.P.A.C.) vs 3-methenylpentane (Nodal).

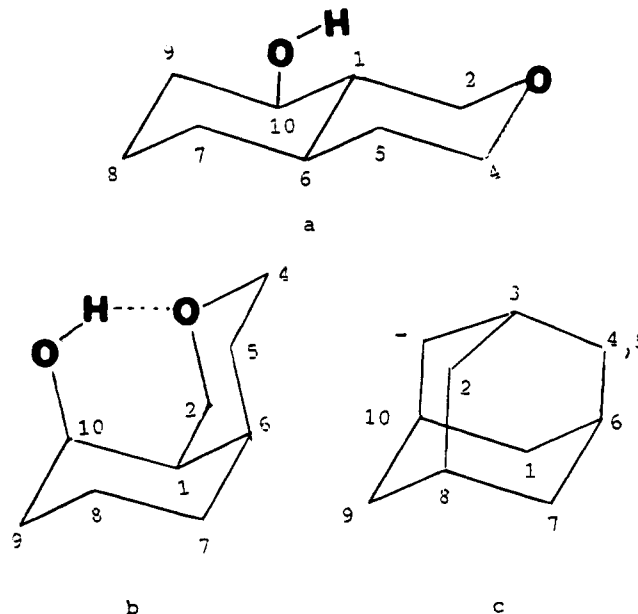


Figure 2. (a) 10-Hydroxy-*trans*-3-oxadecalin, (b) 10-Hydroxy-*cis*-3-oxadecalin, and (c) adamantane (nonstandard locant numbering).

selected as locant no. 1; also, there exist two possible directions in which to traverse the ring, rather than the sole choice of progressing along the chain.

In the case of multiring compounds, the method chosen in most presently used systems is predicated on the von Baeyer system⁵ and begins with projection onto a planar surface and then selecting the smallest set of smallest rings (SSSR)⁶ as the set to be nomenclated. In other words, the guiding principle for all such nomenclatures is predicated on the methods described by Taylor,⁷ which work fairly well for molecules that are topologically described by a two dimensional embedding space.⁸ An important observation is that when Taylor proposed his system in 1948, it was adequate for most of the then-known molecules. An only

[®] Abstract published in *Advance ACS Abstracts*, December 1, 1994.

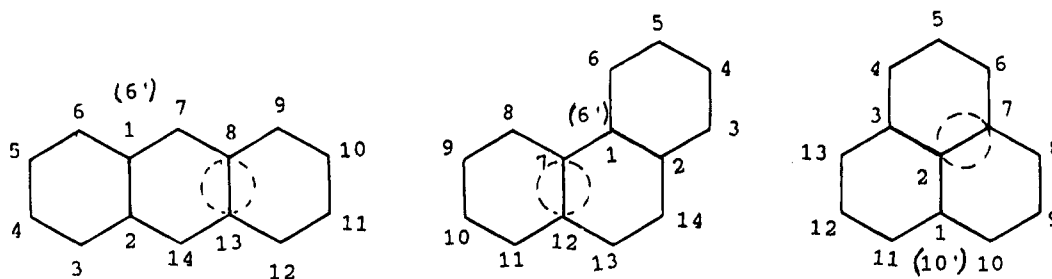


Figure 3. Nonexistence of an Eulerian path for tricyclic unifilar or reticularly-fused compounds.

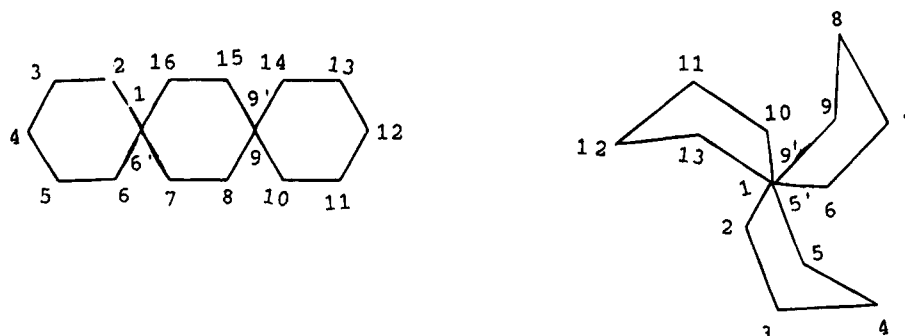


Figure 4. Existence of an Eulerian path (an Eulerian cycle) for spiro-fused tricyclic compounds.

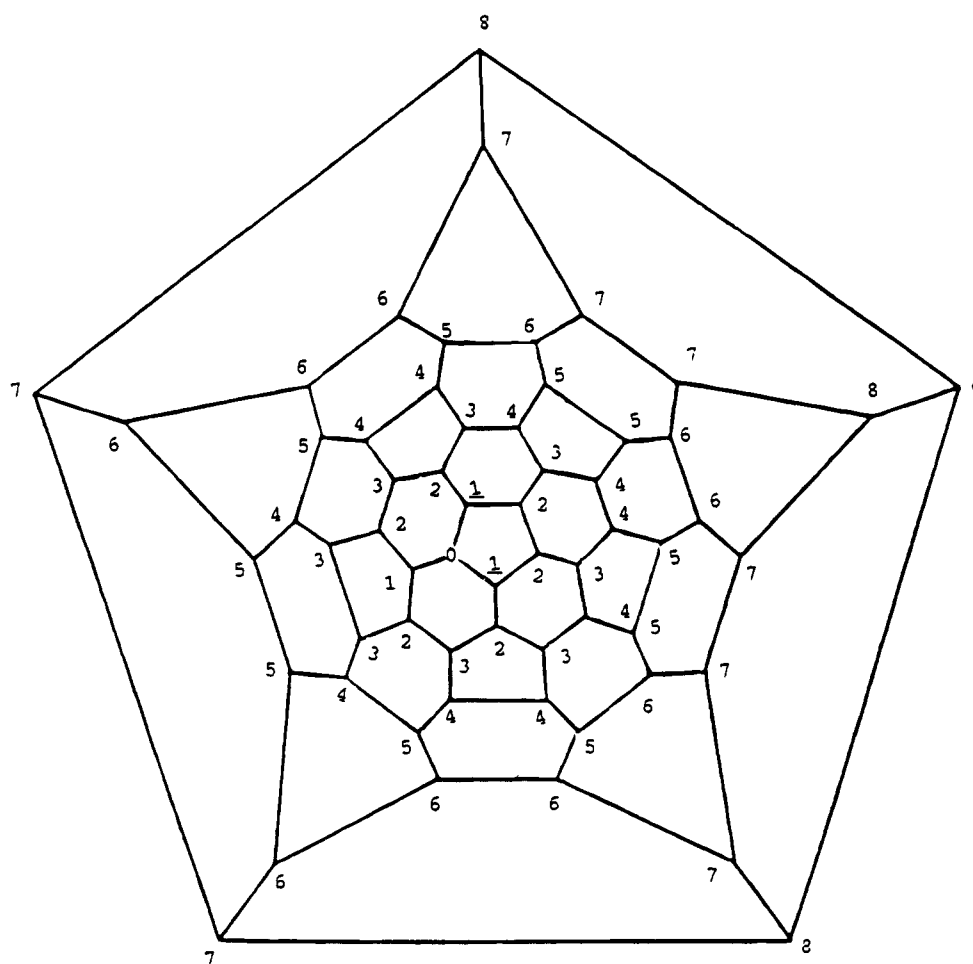


Figure 5. Graph theoretical distances (GTDs) from reference node in C_{60} .

slightly over-simplified explanation would be that chemistry had, at that time, progressed from a prospective of one-dimensional embedding to a two-dimensional one. The next major advance in nomenclature (which paid more than just lip-service to the three dimensional nature of molecules) occurred in 1980, when Goodson⁹ recognized inherent

limitations in Taylor's class of "reticular" compounds and subdivided it into two subclasses, which he called "reticular" and "fisular": The newly-designated class which retained the old name reticular was now limited to those molecules for which there was negligible interaction between atoms of different rings EXCEPT for those atoms common to both

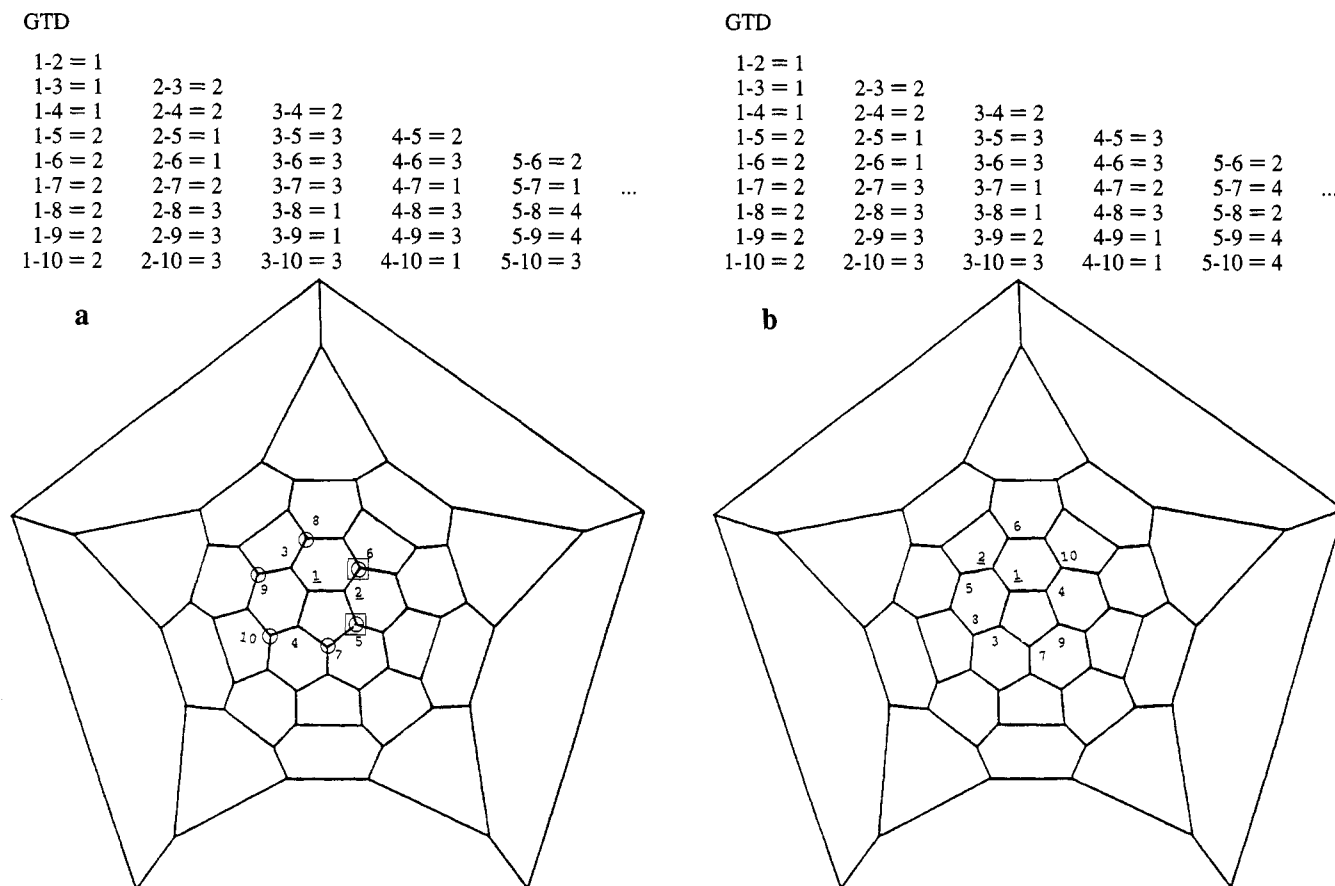


Figure 6. Procedure for formulating canonical numbering of vertices for C_{60} (a) when edge 1-2 is between a hexagon and a pentagon and (b) when edge 1-2 is between two hexagons.

rings. For such ring systems, we could, WITHOUT LOSS, consider projection of the individual rings onto a plane—even though the individual rings might, in the three-dimensional world, be highly puckered. This property is what we had earlier referred to as “essentially coplanar”,¹⁰ and which we illustrated by the *trans* isomer of decalin. Furthermore, we note that even if we were to consider the 3-oxa derivative of *trans*-decalin in which the hydrogen atom in position no. 10 had been replaced by a hydroxyl group (Figure 2a), the oxa atom would still be far enough away from any nonadjacent hydrogen atoms, so that any hydrogen-bonding would be insufficient to approximate an additional ring. For such compounds negligible deviation from “essentially global coplanarity” was the most salient feature. In contradistinction to this, we note a major topological difference in the *cis* stereoisomer of that otherwise same substituted oxadecalin (Figure 2b). If we overlook the distinction between a covalent and a hydrogen bond, the topological differences between this compound and adamantane (Figure 2c) [A nonstandard locant numbering is used in Figure 2c in order to facilitate comparison with Figure 2b.] are relatively insignificant; namely, counting the hydrogen atom as a member of a “ring” there are two more atoms in the “cage” and the connection of the oxa-ring to the all-carbon ring is from locant no. 2 to locant no. 1, rather than to locant no. 8. This is equivalent to saying that the substituted *cis*-decalin is nearly a substructure of an adamantane-type compound and may, similarly, be thought of as possessing an intrinsically three-dimensional character. It is this “essential three dimensionality” that is prototypical of the class designated as *fisular*.⁹ Furthermore, we note that for this *fisular* class

(which includes the fullerenes, fulleranes, etc.), major inadequacies exist with all of the present commonly-used categories of nomenclature¹¹—including the assignment of locant numbers.

2. APPLYING GRAPH THEORY PRINCIPLES TO CHEMICAL NOMENCLATURE

Approaching the problem of assigning an optimal naming for a chemical moiety from a mathematical perspective, one possible scheme begins by considering the graph formed using the carbon atoms as the node set and the bonds between connected carbon atoms as the edge set. We now wish to traverse such a carbon atom skeleton by means of a closed Eulerian path.¹² For example, although we can describe exactly two Eulerian paths for an unbranched acyclic compound and $2n$ (n = number of nodes) Eulerian paths for an unbranched monocyclic compound, the utility of such paths arises primarily when considering polycyclic systems. In particular, we note that no three-ring unifilarly-connected or reticularly-bridged system⁷ contains an Eulerian path (Figure 3); however, if the three rings are spiro-connected,⁷ then an Eulerian path exists (Figure 4). The chemical significance of this is that the vast majority of ring systems containing three or more rings will NOT contain ANY Eulerian paths. HOWEVER, if the graph theory model chosen is to have relevance, then we *must* traverse every edge of the graph in order to completely describe the “electronic” pathway that exists in a molecule. Pragmatically, over the years chemists have tried to circumvent this problem by focusing attention on the union of fused Eulerian paths that cover the entire carbon skeleton and then selecting

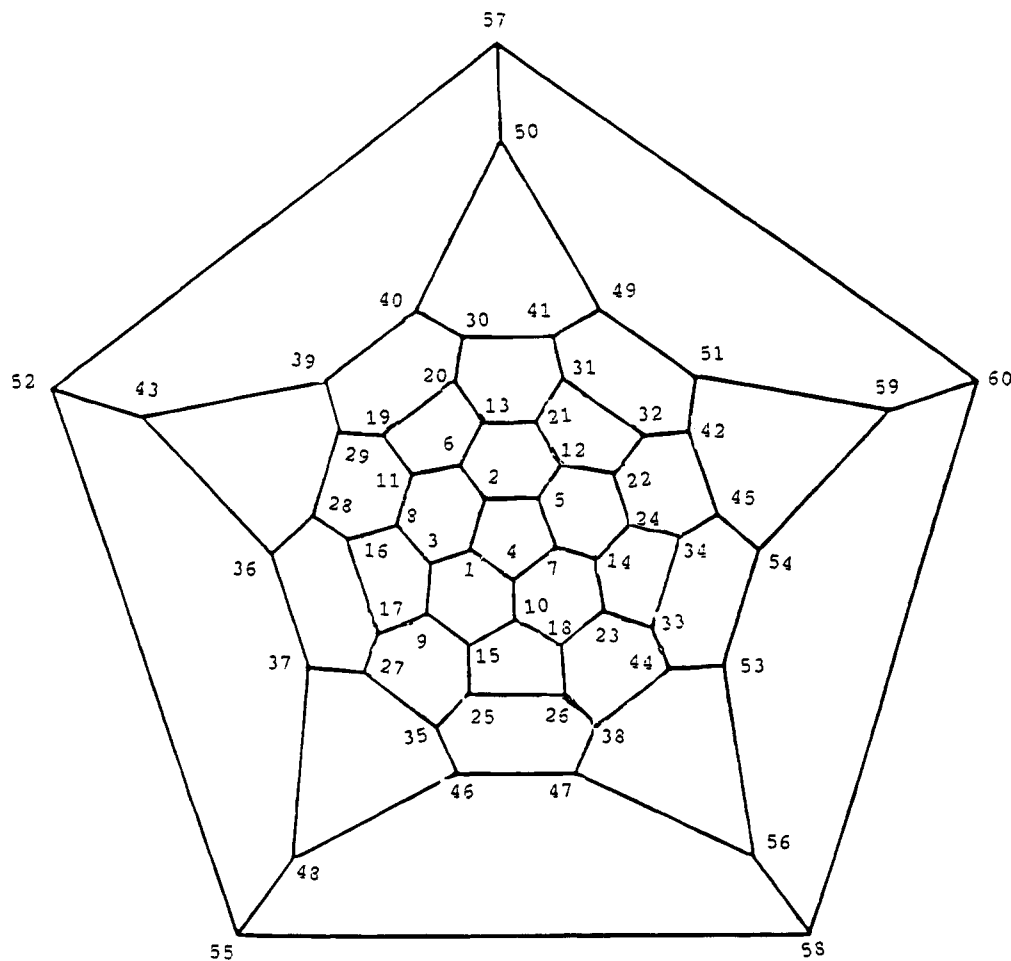


Figure 7. Canonical numbering for all of the vertices of C_{60} .

some minimal set. As was shown in ref 6, such a strategy exacts a price that is relatively inconsequential for molecules that can be well represented using SSSR, i.e., reticular molecules. However, for fisular molecules, the use of SSSR can be damning.

3. A CRITERION AND ALGORITHM FOR NOMENCLATING FISULAR COMPOUNDS

In what was originally an unrelated task, at a 1990 graph theory conference,¹³ we presented an alternative method for assigning names to the vertices of regular (and semiregular) polyhedra with the intent of better characterizing the four Kepler–Poincaré polyhedra. The underlying philosophy of this method was to order the vertices according to increasing graph theoretical distance (GTD) from a selected reference vertex and to so formulate a table of GTDs which gives a canonical ordering of each succeeding vertex of the figure. This was next applied to chemical nomenclature at the 1991 summer American Chemical Society meeting.¹⁵ During the Open Meeting of the Nomenclature Committee of the American Chemical Society on August 22, 1994, the problems of assigning locant numbers to the various fullerenes—as well as to closely related other compounds, such as the fulleranes, etc.—were aired. At this meeting, it was discovered that the existence of ref 16 was not known to this committee. Subsequent conversations and correspondence with Dr. Alan L. Goodson of Chemical Abstract Services and Drs. Reiner Luckenbach and Gustav Penzlin of Beilstein Institute form the basis for extending the locant

numbering from smaller fisular compounds to the class of fullerene, fulleranes, etc. by means of the proposed following algorithm:

1. Select as vertex no. 1, one of the highest atomic number atoms forming the fisular cage. In the case of the C_{60} fullerene (see Schlegel projection¹⁷ in Figure 5) because each vertex is a carbon atom that is at the intersection of two hexagons and one pentagon, all 60 vertices are constitutionally equivalent;¹⁸ consequently, it is completely arbitrary which vertex to designate as no. 1. Furthermore, having arbitrarily selected vertex no. 1, the GTDs to the remaining 59 vertices are shown in Figure 5. In the event that the molecule being numbered is a fullerane, or any other compound having pendant ligands, choose as vertex no. 1 the carbon atom having the highest atomic numbered ligand; however, this is subordinate to heteroatoms in the cage itself.

2. Select as vertex no. 2, one of the vertices that is adjacent to no. 1. In particular, we wish to designate as locant no. 2, that atom which will yield the lowest set of locant numbers for the remaining vertices. For the C_{60} fullerene, we see from Figure 5, that of the three vertices with $GTD = 1$ from vertex no. 1, two are constitutionally equivalent (underlined in Figure 5), but the third is not. [The potential candidates are the one that gives edge 1–2 between a hexagon and a pentagon (Figure 6A) vs one of the two giving edge 1–2 between two hexagons (Figure 6B). Note that for this particular C_{60} model, there is no instance in which two pentagons are adjacent.] Consequently, in order to determine which of these two models we wish to select, we must follow

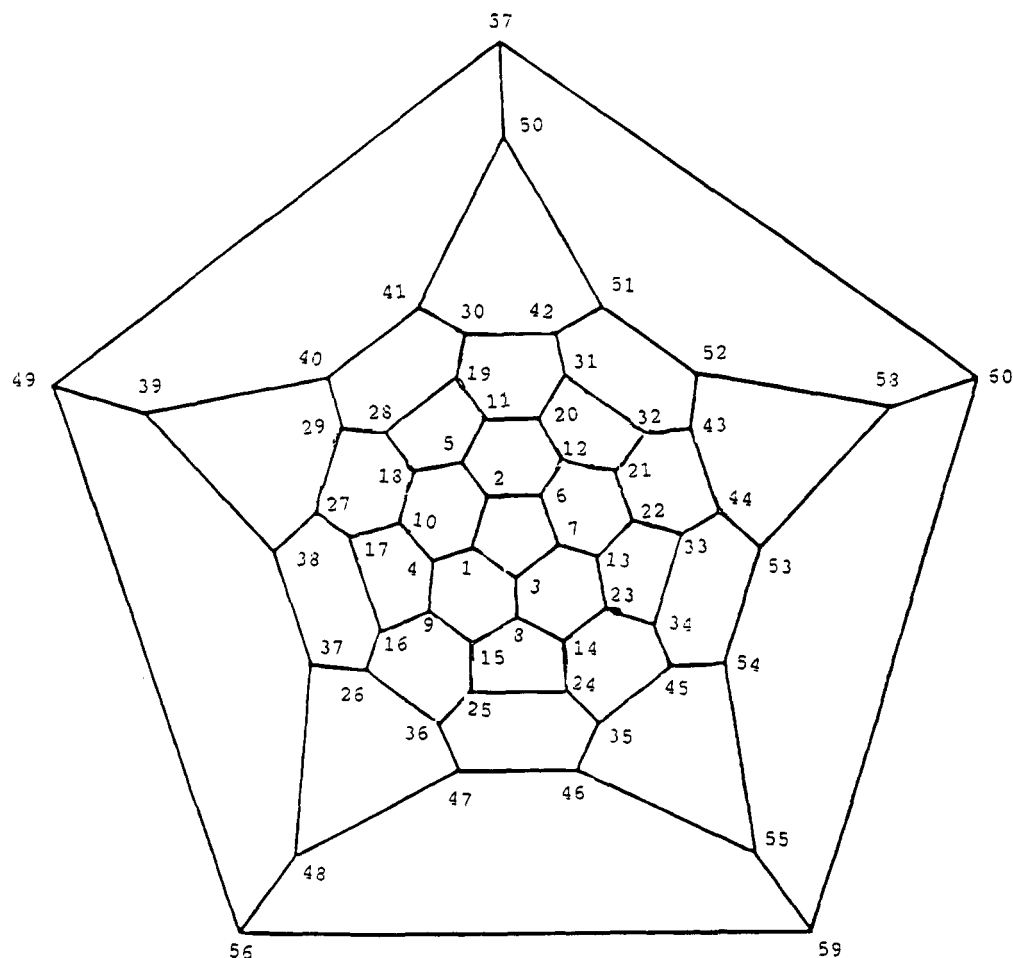


Figure 8. Alternative canonical numbering of vertices for C_{60} —using clockwise assignment to equal GTDs from no. 1.

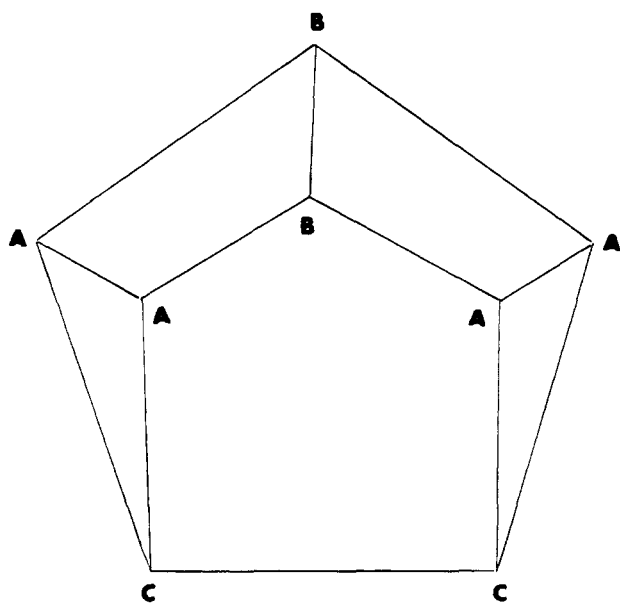


Figure 9. Constitutionally equivalent vertices in cubane.

through each of them—adding successively locant nos. 3, 4, etc. and examining the set of GTD from 1–3, 2–3, 1–4, 2–4, 3–4, etc.

3. From Figure 6A we note that the choice of locant no. 3 from between the two remaining GTD = 1 vertices is arbitrary. However, once locant numbers have been assigned to these two vertices, there is no more arbitrariness in the remaining locant assignments. In other words, although there are six vertices with GTD = 2 from vertex no. 1 (indicated

by a circle), our choice of which vertex to call “5” is limited by first minimizing GTD from vertex no. 2 (indicated by a square) and finally is uniquely determined by minimizing GTD from vertex no. 4. In a similar manner, the remaining 55 vertices are canonically ordered.

Additionally, we note when comparing the two sequences (Figure 6A vs 6B) that, although the GTDs from vertex 1 are the same, from other vertices they may be different. In other words, we compare these two figures by going down the successive columns or across the successive rows until we find the first difference and choose the numerically lowest sequence as the desired one. For columns, this occurs for GTD 2–7: In Figure 6A GTD 2–7 = 2, while in Figure 6B GTD 2–7 = 3. Similarly, when going across rows, this occurs for GTD 4–5: In Figure 6A GTD 4–5 = 2, while in Figure 6B GTD 4–5 = 3. Consequently, by either method, Figure 6A is the one to use in completing the assignment of locants for the remaining 50 nodes. Figure 7, containing the desired locant sequence, is thus generated by ordering next the eight GTD = 3 nodes and then the ten GTD = 4 nodes, etc. (See Figure 5.)

An alternative, in many respects more convenient, locant numbering system is shown in Figure 8. It combines the idea of GTD with clockwise or counterclockwise numbering of nodes with equal GTD. The pragmatic advantage of this scheme is that once the first few locant numbers have been assigned, the remainder can be sequenced without having to make a large number of countings. Furthermore, this alternative scheme suggests a means of distinguishing between the two forms of C_{76} fullerene¹⁹ as well as other

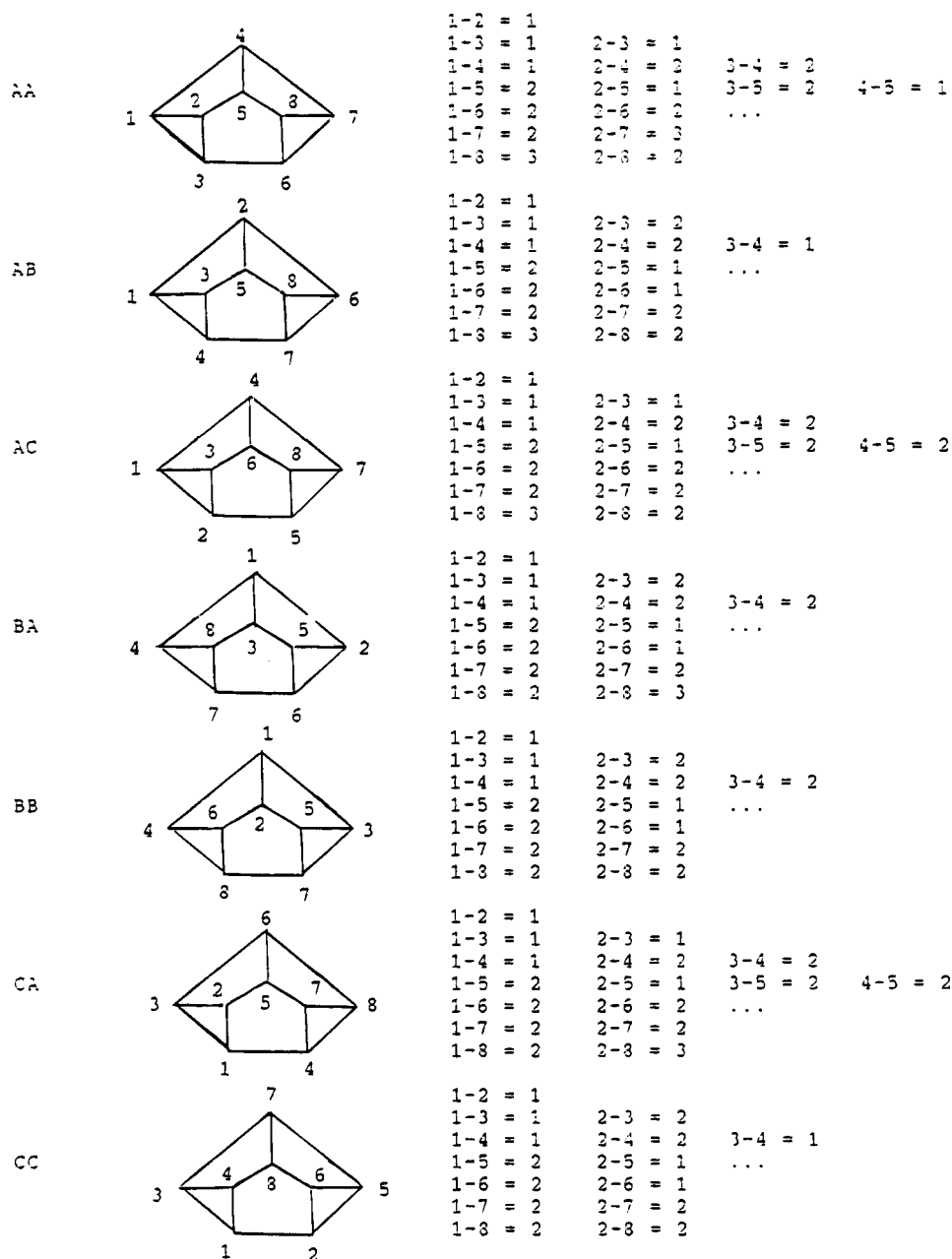


Figure 10. Procedure for formulating canonical numbering for cubane.

such compounds which have a chirality. For such compounds, a "built-in" means of assigning R vs S could be formulated using just the first three GTD = 1 nodes.

4. A FISULAR COMPOUND WITH LESSER SYMMETRY

An additional comment that should be kept in mind is that not all fisular compounds have the extreme symmetry considerations that the five Platonic solids¹⁵ and C₆₀ fullerenes have. Consequently, when choosing which vertex to name as "vertex no. 1", we may have to examine all possible *n* atoms in the cage. For example, in cubane, the eight vertices are of three distinct types (labeled A, B, and C in Figure 9). This means that, in order to get started, we must first examine all possible choices of vertex no. 1 (3), followed by all possible choices of vertex no. 2 (3); i.e., nine potential choices may have to be delineated. In actuality, this number is less since vertices B and C are not adjacent; consequently, combinations BC and CB are not viable, and thus there are only seven combinations that need to be examined. As

mentioned above, there are now two reasonable potential criteria for minimizing the GTDs first from vertex no. 1, then from vertex no. 2, etc.—in order to canonically number the vertices. Unlike the C₆₀ fullerenes, however, in the case of cubane, these two produce different canonical numbering schemes. The numbering schemes are shown in Figure 10 by either going down the columns or across the rows in order to locate the numerically lowest sequence of GTDs. The former of these guarantees that we have the maximum number of adjacent atoms with minimum GTDs. It is also pragmatically simple in that, for our cubane example, we needed only the first column to narrow our focus to the last four (B or C) possibilities, and then only the second column to choose CA as the configuration to use in formulating the canonical numbering. The latter, which would be more efficient for a large number of atoms and which would be necessary for an infinite tessellation of the embedding space, is found by tabulating the GTDs across rows first. For the cubane example, in the assignment of vertices 1–4 we have

the minimum sequence, 1-1,1-1,2,2 for AA, AC, and CA, and a higher sequence for the other possibilities. Continuing to the next row; i.e., thru vertex 5, we see that AA produces the row 2,1,2,1 vs 2,1,2,2 for AC and CA. Consequently, by this criterion, the canonical locant numbering sequence would be the one designated as AA.

5. FUTURE CONSIDERATIONS

It may be noted that this report has focused only on a small, but very important, part of the question of assigning a canonical nomenclature to all potential variations of the class of fisular compounds, namely the numbering of the locants. There are two major reasons for this limited perspective: (1) We foresee potential ambiguities in the designation of R vs S stereoisomers for C₇₆ and higher chiral fullerenes that must still be worked out in order to guarantee consistency and (2) we have not resolved whether there are further logical implications that must be taken into consideration in ascertaining the "superiority" of column vs row ordering in section 3?

REFERENCES AND NOTES

- (1) International Union of Pure and Applied Chemistry, Nomenclature of Organic Chemistry. *Definitive Rules for Section A. Hydrocarbons*; Pergamon Press: Oxford, England, 1979.
- (2) Lozac'h, N.; Goodson, A. L.; Powell, W. H. Nodal Nomenclature—General Principles. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 887–899.
- (3) Elk, S. B. Orismology (The Science of Defining Words) and the Geometrical Foundations of Chemistry—3. What Constitutes a Ring? Presented at the 208th Meeting of the American Chemical Society, August 21, 1994; publication pending.
- (4) Wiswesser, W. J. A Line-Formula Chemical Notation; Crowley: New York, 1954.
- (5) International Union of Pure and Applied Chemistry, Nomenclature of Organic Chemistry. *Definitive Rules for Section A. Hydrocarbons*; Pergamon Press: Oxford, England, 1979; Rule A-31.
- (6) Elk, S. B. Derivation of the Principle of Smallest Set of Smallest Rings from Euler's Polyhedron Equation and a Simplified Technique for Finding This Set. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 203–206.
- (7) Taylor, F. L. Enumerative Nomenclature for Organic Ring Systems. *Ind. Eng. Chem.* **1948**, *40*, 734–738.
- (8) Elk, S. B. Topologically Different Models to be Used as the Basis for Ring Compound Taxonomy. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 17–22.
- (9) Goodson, A. L. Graph-Based Chemical Nomenclature. 2. Incorporation of Graph-Theoretical Principles into Taylor's Nomenclature Proposal. *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 172–176.
- (10) Elk, S. B. Orismology (The Science of Defining Words) and the Geometrical Foundations of Chemistry. 2. Inherent Topology. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 325–330.
- (11) Elk, S. B. Topological Considerations Subtly Inherent in the Formation and Subdivision of Fused vs. Bridged Ring Compounds. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 70–73.
- (12) By an Eulerian path is meant that it is possible to draw one continuous line that traverses the entire network of edges without having to either retrace any edge more than once or to exit and reenter the graph at a different node. Graph theoretically this means that every edge is covered exactly once BUT nodes may be passed through many times. Historically, graph theory is said to have had its beginning when Leonhard Euler discovered that to traverse a set of bridges in the city of Koenigsburg without traversing any bridge twice, either every node of what we now call a graph had to be of even degree or else exactly two nodes were of odd degree—with these two nodes being the first and last nodes of the path. For further discussion of Eulerian graphs see: Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969; pp 64–65.
- (13) Graph Theory Day XX of New York Academy of Sciences; St. John's University: Staten Island, N.Y.; November 10, 1990; proceedings published as ref 14.
- (14) Elk, S. B. Formulation of the Kepler-Poinsot Polyhedra (Including the Degenerate Cases) Using Equal Graph Theoretical Distances to Facet the Vertices of the Regular Polyhedra. *Graph Theory Notes of New York XX* **1991**, 42–49.
- (15) Fourth Chemical Congress of North America and 202nd National Meeting of the American Chemical Society; New York, NY; August 28, 1991; proceedings of the Chemical Information Division program of that date published as ref 16.
- (16) Elk, S. B. Graph Theoretical Algorithm to Canonically Name the Isomers of the Regular Polyhedranes. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 14–22.
- (17) Loeb, A. *Space Structure—Their Harmony and Counterpoint*. Addison-Wesley: Reading, MA, 1976.
- (18) Schubert, W.; Ugi, I. Constitutional Symmetry and Unique Descriptors of Molecules. *J. Am. Chem. Soc.* **1978**, *100*, 37–41.
- (19) Ettl, R.; Ito, C.; Diederich, F.; Whitten, R. L. Isolation of C₇₆, a chiral (D₂) allotrope of Carbon. *Nature* **1991**, *353*, 149–151.