A New Method of Nodal Numbering for Cyclic and Acyclic Structures

Donald J. Polton*

Department of Computer Science, University of Hull, Hull HU6 7RX, England

Received February 4, 1992

Computer studies on the interconversion of certain representations of chemical structures[1] have led to the conclusion that the generally accepted method of basing the enumeration of cyclic and acyclic components of a structure on the largest ring and the longest chain can give rise to problems in the selection of the correct route. These problems are overcome by taking as the primary ring or chain not necessarily that which has the most nodes but that which is the most highly branched. A new system of nodal numbering for cyclic and acyclic systems has been developed which enables codes/names for complex structures to be readily obtained from a connectivity matrix.

INTRODUCTION

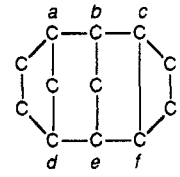
The term "nodal nomenclature" is generally understood to mean the nomenclature system developed by Lozac'h, Goodson, and Powell,[2,3] based on a system of nodal numbering suggested by Lozac'h in 1973. In a nodal numbering system each atom, or node, of a structure is given a unique serial number; these numbers being allocated by a strict, simple set of rules. The advantages of nodal over conventional IUPAC nomenclature have been discussed by Lozac'h.[4]

Nodal numbering has been in existence for many years. In the IUPAC nomenclature[5] certain bridged rings are numbered using an extension of the von Baeyer system (Rules A-31, 32). Since the appearance of the numbering system of Lozac'h, a completely new nomenclature system which is partially based on nodal numbering has been published by Hirayama as the HIRN system.[6] In these systems the numbering of polycycles is based on the largest ring present, and in the Lozac'h and HIRN systems the numbering of branched acyclic structures is based on the longest chain. Another nodal numbering method for polycyclic systems originated by Taylor[7] was used in the notation and nomenclature systems based on the work of Dyson.[8-11] This is based not on the largest ring, but on the smallest set of smallest rings (SSSR), and will not be considered here. Only the system of Lozac'h can however be called a fully nodal nomenclature, although there are even here instances where a node is considered as a substituent rather than as part of the basic system, as with the oxygen of a hydroxyl group.

In the following considerations of nodal numbering, polycyclic and branched acyclic systems are treated separately. Structures without branching nodes are treated similarly in all of the systems mentioned and present no problem. The term "node" is used to indicate an atomic center, which is a carbon atom unless otherwise designated. No attempt is made to progress beyond dealing with the structure as a graph with node and edge variations, i.e., functional groups are not considered at this stage.

Polycyclic Systems. The three methods of nodal numbering of polycyclic systems, von Baeyer, Lozac'h, and HIRN, are all based on the largest ring size present in the polycycle. They differ in the choice between rings of equal maximum size and also in the selection of the primary bridge, and consequently, the primary node and direction of numbering may differ. With both von Baeyer and Lozac'h the longest bridge across the main ring is the determining factor, leading to two possible starting points and four possible circuits of the main ring for each bridge of the same maximum size. The primary difference between these two systems is in the selection between bridges of equal length and in the direction of numbering around the ring. Whereas von Baeyer requires the bridge to bisect the main ring as closely as possible and the numbering to follow the longer path between the two main bridge nodes, Lozac'h selects the primary bridge as that which finishes as closely as possible in the main ring to its starting point and numbers the shortest path first.

The HIRN approach is different. The main ring is numbered so as to give as low a total numbering as possible to the main bridge nodes. Bridge length is only used to select between like enumerations, when that with the longer initial bridge takes precedence. Structure **1** illustrates the difference between the three systems.



1

Here the largest ring is 10-membered, taking in the six bridging nodes *a–f*. There are two of the maximum size 1-node bridges, *ad* and *be*. Of these *be* is selected for the von Baeyer system as it bisects the main ring. For the Lozac'h system *ad* is selected as there are fewer separating nodes (2) than with *be* (4). The structure is symmetrical, hence either bridge node can be taken as the primary node. Taking *a* and *b*, both systems number the main ring anticlockwise, von Baeyer on account of the preference of the 1-node bridge *ad* over the 0-node bridge *cf*, and Lozac'h on account of the proximity of *a* and *d*. HIRN requires the bridge nodes to be given the lowest possible numbers so that *a*, *c*, *d*, or *f* are contenders for primary node, giving the numbering sequence 1,2,3 to bridge nodes. The structure being symmetrical, the choice is between *a* and *c* and is made by selecting the longer bridge as the first bridge, namely, *ad*, and the numbering is clockwise.

To summarize, the main ring numbering of these three systems is shown in Figure 1.

Polycyclic systems such as that of structure **1** present no difficulty in numbering. When a bridge contains a branch

† Present address: 'Rosefarm', Denne Manor Lane, Shottenden, Canterbury, Kent CT4 8JJ, England.

0095-2338/92/1632-0430$03.00/0

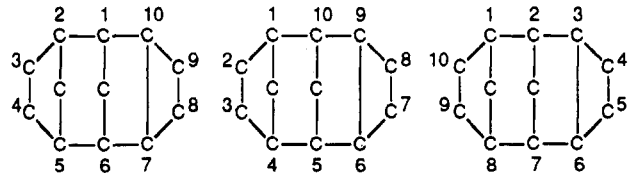
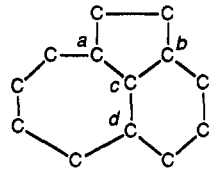


Figure 1. Nodal numbering of the main ring of structure **1** by three methods.

either to the main ring or especially to another bridge, problems are encountered. These problems may not be immediately apparent because, dependent on the way a structure is drawn, it may be quite easy to perceive the correct enumeration visually.

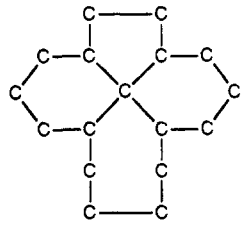
In structure **2** the largest ring is 12-membered and includes



2

all outer nodes. The central node *c* is a branching bridge node. The longest bridge is one-membered, of which there are three to be considered, *ab*, *ad*, and *bd*. For each of these bridges the primary node may be either one of the two bridge nodes, and for each possible primary node, numbering of the main ring may be clockwise or anticlockwise. There are therefore 12 possible enumerations of the main ring. And until all 12 possibilities have been compared it is not possible to number the main ring.

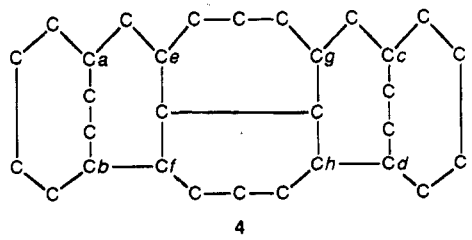
With structure **3**, in which there is a quaternary branching



3

bridge node, the number of possible enumerations increases to 24.

These are relatively simple examples, with only one bridging node. Structures with more than one branch node in the bridge and with additional bridges can give rise to numerous contenders for the correct enumeration. In structure **4** there



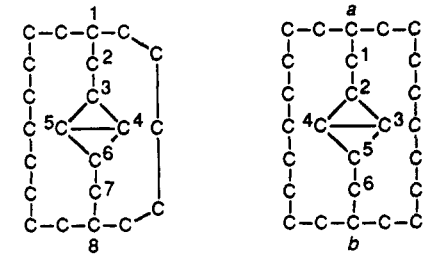
4

are not only the simple 2-node bridges *ab* and *cd* but also the paths *ef*, *eg*, *eh*, *fg*, *fh*, and *gh* to be examined.

The advantages of the HIRN numbering method becomes apparent, using bridge length as a secondary determining factor only. The number of primary contenders for the main ring is greatly reduced. In structure **4** the primary node must be *b* or *d* to give the lowest bridge node enumeration 1-2-6-7.

A greater problem arises when a bridge across the main chain contains a cyclic component.

In structure **5a** the main ring of 17 nodes numbered as

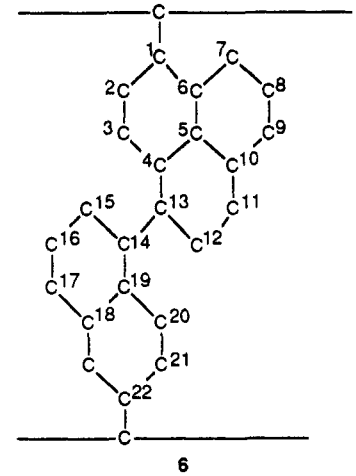


5a, Lozac' h numbering **5b**, bridge enumeration

shown takes the path through the inner bicyclic system 3-4-5-6 (or 3-5-4-6). But if the outer ring is enlarged to 20 nodes, as in **5b**, the bicyclic structure between *a* and *b* becomes part of a bridge.

A distance matrix will show the distance across this bridge (*a* to *b*) as consisting of 5 nodes (1-2-3-5-6 or 1-2-4-5-6), whereas the longest bridge, as required for the nodal systems under consideration, has 6 nodes. This makes it necessary in a case like this to analyze the inner ring system and carry out computer maneuvers to find longer paths through it. In this case two possible paths will be seen: 1-2-3-4-5-6 and 1-2-4-3-5-6.

Again, this example is one of the simplest of its type in that there is only one bridge, and this with few nodes. An even more complex example is the bridge structure **6**, in which a

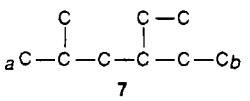


6

bridge of 10 nodes will be found from a distance matrix but which also has 12-, 14-, 16-, 18-, 20-, and 22-node bridges. The 22-node bridge, which takes in all but one node, is required for the nodal numbering systems. The path is shown as 1-22, although the actual enumeration of the systems will differ.

Acyclic Systems. The numberings of acyclic structures in the Lozac'h and HIRN systems are both based on the longest chain. There are differences, which are similar to those for cyclic systems. With Lozac'h, the secondary consideration is the length of the branches on the main chain, priority being given to those which are the longest. With HIRN, the branches must be as near as possible to one end of the main chain, giving as low a numbering of branching nodes on the main chain as possible. Structure **7** illustrates these differences.

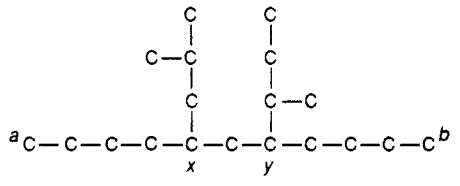
The Lozac'h system numbers from *b* to *a* so as to give as low a locant number to the longer 2-node branch as possible;



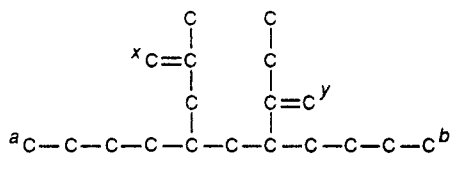
7

HIRN numbers from *a* to *b* so as to give as low a locant to the first branch, irrespective of length, as possible.

With acyclic systems, complex branches may have to be analyzed before the main chain can be numbered or a selection has to be made if there are equal alternatives. Structure **8** is considered as a derivative of the 11-node chain *ab*, and the branches at *x* and *y* have to be compared before the direction of numbering of *ab* can be established. It is worth noting here that, in IUPAC nomenclature, a shorter chain will be given seniority over a longer chain if it has a higher degree of unsaturation. The IUPAC name for structure **9** is based on the 8-node chain *xy* and not the 11-node chain *ab*.



8



9

Gottlieb and Kaplan[12] have suggested a similar approach to nodal numbering, giving precedence to unsaturated chains regardless of their length. Strictly speaking, however, in a nodal numbering system the edges (bonds) should not be considered until the nodal configuration has been established.

In computer processing, similar problems are encountered with acyclic systems as with cyclic systems, entailing analysis and comparison before the correct main chain can be selected and enumerated. This "looking ahead" necessitates lengthy computer procedures which can be eliminated by selecting an alternative main component, be it ring or chain.

A NEW NODAL NUMBERING SYSTEM

To enable a structure to be numbered by the computer from a connectivity matrix with the least difficulty and the simplest programming, a system of nodal numbering has been developed which incorporates as high a degree of branching as possible within the main ring or chain. Size of ring or chain is not considered unless there is more than one of the same size with the same degree of branching.

Again, cyclic and acyclic systems are considered separately.

Polycyclic Systems. Computer studies[1] have shown that the major disadvantage of current nodal numbering systems for polycycles is the complexity of the bridges, and if the system were to give rise to less complex bridges, the process of enumeration would be easier.

To reduce the complexity of the bridges, the complexity of the main ring must be increased, that is, as many branching nodes as possible must be transferred from the bridge systems to the main ring. This is done by taking as the main ring that which has the greatest number of bridges, irrespective of its size. If there happens to be a complex fused-ring system embedded within a very large cycle, this means striking a path through this system rather than taking it as a bridge. In structure **5b** this path is the same as that taken by current nodal numbering systems in **5a**.

A further disadvantage of current nodal numbering is the use of the bridge of the greatest length as the secondary deciding factor in selecting from alternative bridges. If the shortest, rather than the longest, bridges are chosen as deciding factors, these can be readily obtained from a distance matrix. There may be several of the same shortest length, in fact there is more likely to be a multiple of short than long bridges. Selection between equal length alternatives is made by taking those which finish as close as possible to their starting point, as with the method of Lozac'h. For each bridge contender there are now four directions of numbering, two from each bridge node and two in each direction. The primary node and direction of numbering is decided by ensuring that branch nodes have as low a total enumeration as possible, as with HIRN.

Having found the primary bridge, determined from which node to start and in which direction to number the ring, and numbered the main ring, bridges are quoted in increasing order of length, so that the bridges likely to be the most complex are dealt with later. A branching bridge, when numbered, will sometimes reveal a new shorter bridge, so that a strict sequence of increasing length is not maintained throughout. When a branching bridge is numbered, the branch nodes in the bridge are given the next available serial node numbers, the numbering being from the lower to the higher locant. Another bridge numbering method has been considered in which bridges are selected in order of node position. In this case the bridges will appear in strict ascending order of primary locant, starting with the principal shortest bridge. Software has been written to produce enumeration codes according to both of these methods, but only the first of these will be considered from this point.

The algorithm for analysis of a polycycle is as follows:

(1) Find the largest ring(s) with the most branching nodes.

(2) Make an adjacency matrix using only those main ring nodes which are branched and those nodes not in the ring.

(3) Distinguish between outer ring nodes and bridge nodes in the adjacency matrix.

(4) Modify the adjacency matrix by zeroing any connections between adjacent main ring nodes.

(5) Create a distance matrix from the reduced adjacency matrix. The distances will be one more than the number of nodes in the bridge, i.e., the number of edges (bonds). In this way 0-node bridges are not confused with unbridged node pairs.

(6) Read from the ring node section of the distance matrix the shortest path between ring nodes. On the first run these are contenders for the main bridge.

(7) Select from (6) the bridge(s) which finish as close as possible to the primary node.

(8) Number the ring both ways from both ends of the primary bridge and reject longer paths. There will now be two possible paths for each shortest bridge (or four for symmetrical bridges).

(9) For direction, select the circuit which gives the lowest bridge node locant at the lowest position.

(10) Give precedence to a node with the greatest number of connections (3 or 4), i.e., quaternary if there is one.

(11) Number any nodes in the bridge from lower to higher numbered ring node. Check the number of

bridges found against the minimum ring count (SSSR count) minus 1. If the numbers are equal, the process is finished.

(12) Trace and remove from the adjacency matrix all of the used bridge bonds.

(13) Relabel any newly numbered branching nodes from the bridge as ring nodes.

(14) Create a new distance matrix from the reduced adjacency matrix.

(15) Return to (6) and continue until all bridges are found.

Taking structure **5b**, the main ring would be smaller than the largest ring, consisting of 17 rather than 20 nodes. But it has six rather than merely two branch points. Following the procedure above, the first four nodes of the structure would be numbers 2, 3, 4, and 5 (the shortest bridge is 0-node), and numbering continues around either of the outer branches (the whole structure is symmetrical).

Description of Numbering System for Polycycles. In current descriptions of nodal numbering systems the bridge node locants are written in superscript form. Using the new system the enumeration the bridges of structure **5b** are, in the Lozac'h format bridge length$^{\text{locant comma locant}}$

$$0^{1,3} \qquad 0^{2,4} \qquad 9^{6,16}$$

whereas the nodal system of Lozac'h gives

$$6^{1,11} \qquad 0^{22,24} \qquad 0^{23,25}$$

For the purposes of computer programming, superscripts are best avoided as their presence requires introduction of case shift characters and translation by the computer every time the structure is read. Instead, a normal case character is used to separate the three numbers describing each bridge. Moreover, these three numbers are placed in their logical order (first locant)–(bridge length)–(second locant). Using "–" as a separator between node descriptors and "," between bridge descriptors, the new description of the bridge system of structure **5b** becomes

1–0–3,2–0–4,6–9–16

and prefixing it by the main ring size followed by a separator, here "..", the full notation becomes

17..1–0–3,2–0–4,6–9–16

Of course, the characters used as separators are arbitrarily chosen. Those used are selected as they give a neat and readable appearance. The double stop is used after the main ring size to show that the system is cyclic, as opposed to a single stop for acyclics. This compares with the main ring prefix 0 used by Lozac'h to show a cyclic system. In the case of a monocycle, the separator ".." is not followed by bridge information and merely indicates a ring rather than a chain.

The part structure **6**, if included in the main ring, would follow the path shown 1–2–3–..–22. This would take in all branch nodes, leaving only one node unnumbered. If the outer ring is of 50 nodes, with each section between the two outer bridge nodes equal at 24 nodes, the coded structure is

48..1–0–6,4–0–13,5–0–10,14–0–19,18–1–22,23–24–48

and the unnumbered node, between nodes 18 and 22, has the number 49.

The reverse numbering, 22 to 1, gives a closer end point to the first bridge, but the bridge is longer (1-1-5) and is therefore rejected.

If this were a bridge rather than part of the main ring, a short 10-node path such as 1-2-3-4-13-14-19-20-21-22 would be chosen, which would be obtained from a distance matrix.

To name this ring using current nodal numbering systems would entail lengthy bridge analysis.

Alternative Main Ring. In this description of the new system, the main ring has been taken as the largest with the most branching nodes. Equally, the main ring could be the smallest with the most branching nodes, in which case the course selected in structure **6** would include the 14-node chain

1-6-7-8-9-10-5-4-13-14-19-18-C-22

which also uses all of the branching nodes 1, 4, 5, 6, 10, 13, 14, 18, 19, and 22.

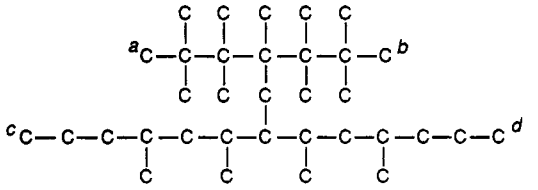
However, although this method looks like a viable alternative to the first suggestion, with the much longer bridges a more detailed analysis is likely to be required if, after establishment of the main ring, there still exists a complex branching bridge system.

Finally, it should be noted that the number of bridges is the same however the system is numbered. It is the number of rings (the SSSR count) in the system minus 1 (the main ring). In bridge detection by the computer, a count is kept to ensure accuracy and to show when all bridges have been found.

Enumeration codes of the above type for cyclic compounds have been produced using Turbo Pascal on the Opus PCIV computer.

Acyclic Structures. For acyclic systems the principal path through the structure is that with the greatest number of branches. Selection is made by computing a chain value which is the sum of the nodal values of 2 for a quaternary node and 1 for a ternary. The consequence of this is that a chain with two quaternary branching nodes is preferred to one with three ternary branching nodes.

In structure **10** the longest chain is from *c* to *d*, 13 nodes.

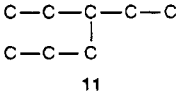


10

Five of these nodes are branching, and the chain value of *cd* is 5. This is normally taken as the main chain in systems of nomenclature and notation, leaving the complex branched system *ab* as a substituent.

There are four chains with the greatest number of branching nodes, from *a* and *b* to *c* and *d*. Each of these has 12 nodes, of which six are branching. With three quaternary nodes their total chain value is 9. The chain from *a* to *b* has only five branching nodes but they are all quaternary. Therefore, its chain value is 10, and *ab* is selected as the main chain. Of course the total nodal value as counted by the computer is the same as the number of branches.

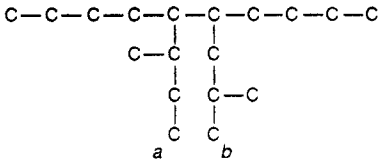
Only when there is a choice of chains with the same number of branches does the length of the chain come into effect. Then the longer chain is taken, leaving as few nodes as possible to deal with later. In structure **11** the main chain is a 6- rather than the 5-node chain.



11

The main chain is numbered starting from the end which is closest to a branching node. If there is a choice of ternary or quaternary node, then the numbering starts from the end nearest the quaternary. If both directions are similar at this stage, then the next nearest branching nodes, if any, are considered. In structure **10** the directions *a* to *b* and *b* to *a* are equivalent.

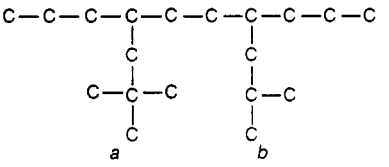
In structure **12**, the main chain is between *a* and *b* with



12

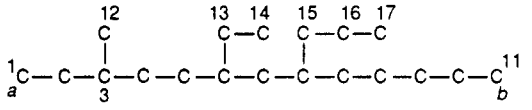
eight nodes and four branches, the longest chain of 10 nodes only having two branches. Since the terminal node *b* is nearer to a branch than the terminal node *a*, the numbering proceeds from *b* to *a*.

In the structure **13** the numbering goes from *a* to *b* because of the quaternary node next to the chain end *a* being given precedence over the ternary node near end *b*.



13

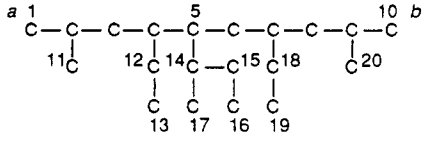
Once the main chain numbering has been established, the branches are numbered strictly in the order in which they occur from the beginning of the main chain. There is no priority for longer chains. This is a logical approach for computer processing as there is no need to examine chain lengths first and decide what to do if there are identical ones. It also simplifies the presentation of the final code. If there are unequal length branches to choose from, the shorter is chosen as the primary branch. In structure **14** the single node branch on main chain node 3 is numbered before the longer branches.



14

The main chain is numbered from *a* to *b*, nodes 1 to 11, and the branch nodes are numbered as shown.

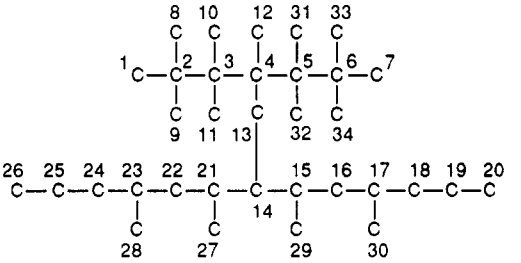
Branched substituents are numbered in their entirety before going on to the next branch along the chain, even though this may be unbranched. Complex branches are treated in the same way as an entire structure, the longest route with the most branches being followed, except that the numbering must start from the node linked to the parent chain. Structure **15** illustrates this.



15

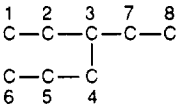
In fact, highly branched substituents will be a rarity, as in the proposed numbering system the branching nodes will have been incorporated into the main chain in most cases. Only when there are more than two highly branched areas in the structure will the case of a highly branched substituent arise.

In structure **15** the direction of numbering of the main chain is decided by the position of the branching substituent. This is nearer to *a* than *b*, and all other equal-length substituent positions being the same distance from both *a* and *b*, the enumeration is as shown. The single-node branch nearest *a* becomes number 11, and the node attached to node 5 becomes number 14. Now this branch must be analyzed before the following 2- and 1-node branches can be numbered. The longest branch is obviously visually seen to be three-membered, but this has to be worked out by studying the lengths of the



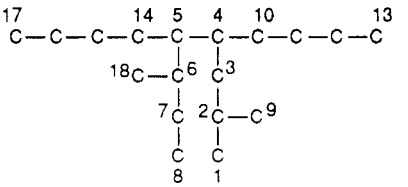
7.2-1,2-1.3-1,3-1,4-1,4-8,14-6,21-1,23-1,15-1,17-1,5-1,5-1,6-1,6-1

10



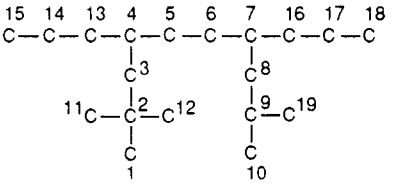
6.3-2

11



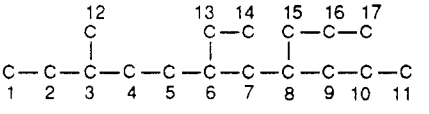
8.2-1,4-4,5-4,6-1

12



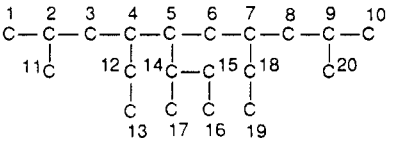
10.2-1,2-1,4-3,7-3,9-1

13



11.3-1,6-2,8-3

14



10.2-1,4-2,5-3,14-1,7-2,9-1

15

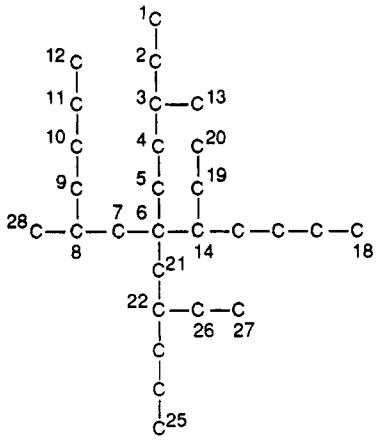
substituent and its branch, and their positional location with respect to one another. The input source could have had this branched branch described in any manner. Having established the main branch numbering, its side chain is then numbered immediately afterwards. The result is as shown.

Describing the Structure. Superscript characters are best avoided for the same reason as with cyclic structures. They are merely a nuisance in computing, having to be translated to and from something else in order to process them. The structure is described much in the same way as for cyclic structures, by quoting the length of the main chain and following it by its substituents in strict ascending order of location. A single stop is used as a separator between the main chain and its branches, and the branches are separated by commas. When a branching substituent is encountered, this is fully numbered before continuing further along the main chain. The format used is

main chain length · [branch position – branch length,]

The codes for fully enumerated structures **10–15** are given in the format presently used on the previous page.

Structure **16** is a further example, in which there are initially three contenders for main chain:



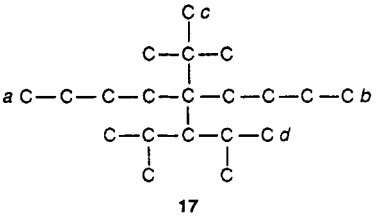
12.3–1,6–5,14–2.6–5,22–2,8–1

16

In this case there are three five-membered chains on the same node. The senior chain contains the one which leads to the lowest numbered side branch.

To detect the main chain visually using this system, it is only necessary to redraw the structure with nonbranching nodes replaced by chain length, when the path of branching nodes is readily seen.

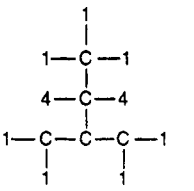
In structure **17**, the longest chain is the nine-membered



17

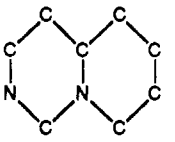
chain *a–b*. The most highly branched chain, from *c* to *d*, has six nodes, four of which are branched. Of these, two have quaternary branching and two have ternary branching. Redrawing this structure as in **17a**, the main chain is easily seen.

Heterocyclic Atoms and Unsaturation. Node and edge variation, i.e., heterocyclic atoms and unsaturation, present no problems. They can be suffixed to the main code. For



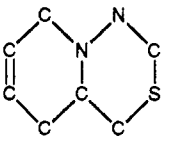
17a

computer studies they are separated from each other and from the main system by a colon. Atomic symbols are placed in alphabetical order (although atomic number order would perhaps be better), and their locants are separated by commas. For double and triple bonds E and Y are used (-ene and -yne), followed by their locants. E without a locant implies aromaticity. Structures **18–20** illustrate this usage.



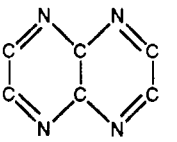
10..1–0–6:N1,3

18



10..1–0–6:N1,2S4:E8

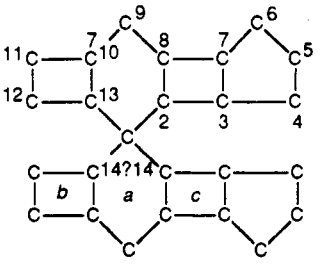
19



10..1–0–6:N2,5,7,10:E

20

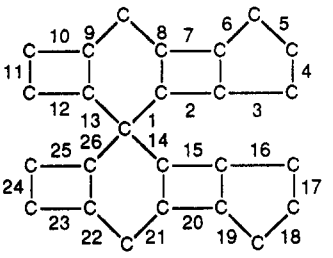
Spiro Structures. With the suggested nodal numbering system, there remains one problem, that of complex spiro structures. The idea behind the new system is to prevent having to build numerous alternative sequences to compare before the primary numbering can be completed. With spiro structures, there are two possible directions of numbering at each spiro node. If the spiro node is embedded in a polycyclic system, there is no problem. It is treated as a quaternary node and will either be incorporated into the main ring or it will be dealt with in the enumeration of the bridges. But if the spiro node alone connects two ring systems, there will be two possible directions of numbering at the spiro connection, and the correct route cannot be determined until both routes around the spiro connection have been studied. In the case of a spiro connection leading to a fused-ring system there may be a build up of possible routes as in structure **21**. Here the



21, nodal numbering

enumeration of the ring system *a* cannot be decided until it has been determined which of the 4-node rings *b* or *c* is to be numbered first. This problem could be overcome by treating spiro connections not as ring fusions but as separate ring systems, as is the case with some systems of nomenclature.

Another idea, which has not been worked upon, is to create a system of "edge numbering" rather than nodal numbering. The bonds rather than the nodes are allocated serial numbers, and in a spiro connection the route will go through the spiro node twice. The route will have to take in as many branching nodes as possible, and there will be an opportunity to consider unsaturation and stereo configuration as deciding factors in the choice of circuit. Structure **21a** shows how this type of enumeration would work. The structure coding is no problem, bonds not in the circuit would be represented by the sequence



21a, edge numbering

(end of bond 1)–(length of connection)–(start of bond 2) and two bonds which meet at a spiro node would have a code such as (bond 1)–S–(bond 2).

A possible edge-numbered code for structure **21a** is 26.1–0–8,2–0–7,9–0–13,1/26–S–14,14–0–21,15–0–20,22–0–26.

CONCLUSION

This paper is concerned only with the nodal configuration of cyclic and acyclic structures and is not intended to present a fully descriptive system of numbering for all chemical structures. Its purpose is to show the drawbacks of existing systems and to point out the avenues open for investigation.

The suggested coding is only one of a number of possibilities based on taking the principal component as that which is the most highly bridged or branched. There is probably a future for nodal nomenclature, and if it is to be eventually taken into general use, all avenues of thought have to be considered. If a nodal nomenclature is ever adopted, it may have to run side by side with conventional nomenclature, and with evermore increasing use and dependence on computer methods, the simplicity of the nodal nomenclature will eventually show its superiority. Perhaps the present-day naming of chemical compounds will at some distant time look as old-fashioned and peculiar as the symbols used by early alchemists do now.

ACKNOWLEDGMENT

I acknowledge the contribution of Dr. G. H. Kirby of the Department of Computer Science, Hull University, under whose supervision the studies of which this paper forms a part were carried out.

REFERENCES AND NOTES

(1) This work is part of a postgraduate research study (Polton, D. J. Ph.D. Thesis, University of Hull, 1991) which is to be published in revised form: Polton, D. J. *Chemical Nomenclatures and the Computer*; Research Studies Press Limited: Taunton, Somerset, England, in press.
(2) Lozac'h, N.; Goodson, A. L.; Powell, W. H. Nodal Nomenclature—General Principles. *Angew. Chem., Int. Ed. Engl.* **1979**, *18*, 887–899.
(3) Lozac'h, N.; Goodson, A. L. Nodal Nomenclature II—Specific Nomenclature for Parent Hydrides, Free Radicals, Ions and Substituents. *Angew. Chem., Int. Ed. Engl.* **1984**, *23*, 33–46.
(4) Lozac'h, N. Principles for the Continuing Development of Organic Nomenclature. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 180–185.
(5) International Union of Pure and Applied Chemistry. *Nomenclature of Organic Chemistry, Sections A–F* and *H*; Pergamon: Oxford, 1979.
(6) Hirayama, K. *The HIRN System, Nomenclature of Organic Chemistry*; Maruzen: Tokyo, Springer-Verlag: Berlin, 1984.
(7) Taylor, F. L. Enumerative Nomenclature for Organic Ring Systems. *Ind. Eng. Chem.* **1948**, *40*, 734–73.
(8) Dyson, G. M. *A New Notation and enumeration system for Organic Compounds*; Longmans Green and Co.: London, 1947.
(9) *Rules for IUPAC Notation for Organic Compounds*; Longmans Green and Co.: London, 1961.
(10) Dyson, G. M. Some New Concepts in Organic Chemical Nomenclature, unpublished.
(11) Polton, D. J. Conversion of the IUPAC Notation into a Form for Computer Processing. *Inf. Storage Retr.* **1969**, *5*, 7–25.
(12) Gottlieb, O. R.; Kaplan, M. A. C. Replacement-Nodal-Subtractive Nomenclature and Codes of Chemical Compounds. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 1–3.