

- (65) Zykov, A. A. *Theory of finite graphs*; Nauka: Novosibirsk, 1969.
- (66) Syslo, M. M. Fundamental sets of cycles of a graph. *Zastov. Mat.* 1973, 13, 399-409.
- (67) Hubicka, E.; Syslo, M. M. Minimal bases of a graph. *Recent advances in graph theory*, Proceedings of 2nd Czechoslovak Symposium on Graph Theory; Fielder, M., Ed.; Academia: Prague, 1975; pp 283-293.
- (68) Syslo, M. M. On cycle bases of a graph. *Networks* 1979, 9, 123-132.
- (69) Syslo, M. M. Minimum length cycle bases of a graph. *Methods Oper. Res.* 1980, 37, 285-290.
- (70) Kolasinska, E. On a minimum cycle basis of a graph. *Zastov. Mat.* 1980, 16, 631-639.
- (71) Tarjan, R. Depth-first search and linear graph algorithms. *SIAM J. Comput.* 1972, 1(2), 146-160.
- (72) Deo, N. Breadth- and depth-first searches in graph-theoretic algorithms. *J. Comput. Soc. India* 1974, 4(2), 19, 31-37.
- (73) Horton, J. D. A polynomial-time algorithm to find the shortest cycle basis of a graph. *SIAM J. Comput.* 1987, 16(2), 358-366.
- (74) Hendrickson, J. B.; Grier, D. L.; Toczko, A. G. Condensed Structure Identification and Ring Perception. *J. Chem. Inf. Comput. Sci.* 1984, 24, 195-203.
- (75) Hendrickson, J. B.; Toczko, A. G. Unique Numbering and Cataloging of Molecular Structures. *J. Chem. Inf. Comput. Sci.* 1981, 23, 171-177.
- (76) 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, 204-206.
- (77) Elk, S. B. Effect of Taxonomy Class and Spanning Set on Identifying and Counting Rings in a Compound. *J. Chem. Inf. Comput. Sci.* 1985, 25, 11-16.
- (78) 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.

## Theoretical Aspects of Ring Perception and Development of the Extended Set of Smallest Rings Concept

GEOFFREY M. DOWNS, VALERIE J. GILLET, JOHN D. HOLLIDAY, and MICHAEL F. LYNCH\*

Department of Information Studies, University of Sheffield, Sheffield S10 2TN, U.K.

Received October 18, 1988

There are many unresolved issues concerning the definition of an optimum ring set for retrieval purposes. This paper considers the problems associated with processing planar (two-dimensional) representations of three-dimensional structures. To overcome the ambiguity of such representations, a new ring set is defined in terms of simple faces and cut faces. The concept of a cut-vertex graph is introduced to explain the combinatorial relationship between the number of simple faces and the number of planar embeddings.

### INTRODUCTION

The previous paper<sup>1</sup> reviewed algorithms and ring sets devised to represent and handle chemical structures and reached the conclusion that none of them presents an ideal solution even for specific structures. It is interesting to note that many of the methods developed have resorted to the use of heuristics, indicative of an underlying and hitherto unresolved problem.

The work on ring perception reported in this series of papers has concentrated on providing a solution for generic environments, of which the specific environment is a special case. In particular, it has concentrated on extension of ring perception capabilities into structurally explicit generics, i.e., those described in full structural detail represented by partial connection tables. To provide a general solution, it has been necessary to consider the theoretical basis of ring perception in greater detail. Due to the lack of suitable formalisms for the description of generalized ring systems as a whole, it has been necessary to consider individual cases and extrapolate to reveal the underlying theory necessary to support such formalisms. This paper presents the conclusions of these considerations, which are reported more fully in reference 2. The task has not proved trivial, but as Elk<sup>3</sup> states

"What is important is the recognition of the problem rather than an attempt to 'paper over' the inconsistencies. *Ad hoc* solutions would never have been selected in the past if a 'simple', 'obvious' solution existed."

As shown in the preceding review paper,<sup>1</sup> Zamora<sup>4</sup> has achieved a certain degree of generalization by defining three distinct classes of ring system in terms of overlap within the vertex and edge sets of the smallest set of smallest rings (SSSR). Methods that reduce a graph to its basic form, such as those of Lederberg,<sup>5</sup> Carhart,<sup>6</sup> Sridharan,<sup>7</sup> and Balaban<sup>8</sup> similarly go only part way to providing a generalization and do not help particularly in making a solution any easier.

Analyses of the distribution and type of ring systems within specific structure databases, by Adamson et al.,<sup>9,10</sup> have shown

that most occurrences are of very simple systems. The more recent studies of the project's database of structurally explicit generics, by Mawby<sup>11</sup> and Kirk,<sup>12</sup> show similar results. However, ring perception techniques have to cater for the worst possible cases of complex systems, and they are judged by their handling of these cases. The review paper has shown that current techniques have difficulty with the more complex cases, and many instances of failure have been mentioned. Before extension to the complexity of variation allowed in structurally explicit generics, it seemed sensible first to develop a sound perception technique that would process the worst-case specifics, rather than build on the recognized limitations of current specific techniques. It should be noted, therefore, that most of the circumstances considered in this research are extreme cases and are not necessarily representative to the types of ring system commonly encountered. Most ring systems in specific structures are simple to process and do not cause problems for other techniques nor for the extended set of smallest rings (ESSR) concept presented here.

### ASPECTS OF GRAPH THEORY

This paper assumes an elementary knowledge of graph theory; further explanation, clarification, and definition of the concepts outlined below are given in the many general treatments and introductions to the subject, such as those by Behzad and Chartrand,<sup>13,14</sup> Berg,<sup>15</sup> and Mayeda,<sup>16</sup> and the classic work by Harary.<sup>17</sup> More applied works include those by Marshall<sup>18</sup> and Bondy and Murty<sup>19</sup> and the excellent presentation by Deo.<sup>20</sup> A few texts deal specifically with chemical applications, including those by Balaban<sup>21</sup> and the highly recommended two-volume work by Trinajstić.<sup>22</sup> This paper also uses the terms and definitions given in the review paper<sup>1</sup> such as component, block, simple cycle, fundamental basis,  $\mathcal{H}$ -ring and  $\beta$ -ring.

The long-recognized correspondence of the atoms and bonds of a chemical structure diagram to the vertices and edges of a graph allows graph theory to be used to represent, manip-

ulate, and analyze chemical structures. These graphs are unoriented and without self-loops or multiple edges.

It is evident that visual inspection to determine the rings in a structure depends upon the way a three-dimensional chemical structure is projected onto a two-dimensional plane. This has major implications for defining a ring set and stems from the problems caused by moving down a dimension. The process of mapping  $n$  dimensions to  $n - 1$  dimensions is referred to as **embedding**. If the graph representation of a three-dimensional structure can be embedded in a plane such that no two edges intersect other than at a vertex, then the graph is said to be a **topological planar graph**,  $G$ . Such a planar graph divides the plane into **regions**. A region is characterized by the set of edges,  $e$ , and vertices,  $v$ , forming its boundary and is not defined in a nonplanar graph, or even in a planar graph not embedded in a plane (i.e., it is a property of a particular plane representation of a graph). Any two points in the region can be connected by a continuous curve that meets no vertices or edges. A region of a planar graph corresponds to a **face** of the equivalent three-dimensional structure. The boundary of a region is the set of all edges that touch that region. The **contour** of a region is defined as the cycle formed with the edges of the boundary of the region that contains the region as its interior.

From Jordan's curve theorem (after Jordan<sup>23</sup>) it can be shown that embedding in a plane results in exactly one unbounded region, otherwise called the **infinite region**, which has no contour. All other regions are bounded, have exactly one contour, and are called **finite regions**. The contours of the different finite regions constitute a cycle basis (i.e., the minimum number of linearly independent cycles covering all vertices and edges) with their number (the nullity,  $\mu$ ) predicted by the Cauchy formula,  $\mu = e - v + 1$  (after Cauchy<sup>24</sup>). Note that for a simple polyhedron the total number of regions or faces,  $f$ , including the infinite region or face, is given by Euler's polyhedral formula,  $v - e + f = 2$  (after Euler<sup>25</sup>), i.e., one more than the Cauchy formula.

A planar graph may be embedded in a plane to produce a Schlegel diagram (after Schlegel<sup>26</sup>), such that any region of specified edges can be made into the infinite region. Obtaining some degree of uniformity by embedding so that the infinite region is as large as possible (the **maximal Schlegel projection**) is a task much discussed by Elk.<sup>27</sup>

It is often easier to envisage embedding upon a sphere to understand the topological significance of the infinite region. A graph can be embedded in the surface of a sphere if and only if it can be embedded in a plane, and so the two representations are directly comparable. In terms of regions on the sphere, it is immediately obvious that there is no real difference between finite regions and the infinite region. A plane may thus be regarded as the surface of a sphere of infinitely large radius. If it is not readily obvious whether or not a structure can be embedded on the surface of a sphere, imagine an elastic model of the structure with a deflated balloon in the center. Inflate the balloon and the polyhedron will gradually flatten and stretch onto the outer surface of the balloon. This process flattens out any surface irregularities of the original three-dimensional model; i.e., it reduces any inherent three-dimensionality to two dimensions by turning any nonconvex areas to convex.

If the spherical representations of two projections of a planar graph are rotated and, if necessary, the regions distorted without letting a vertex cross an edge, then they may become identical. If such manipulation shows all possible projections to be identical, then that graph has a **unique** embedding in a plane or sphere. For instance, the projections i and ii of Figure 1 are of the same graph but rotation will not make them coincide. The first projection has no regions bounded by five

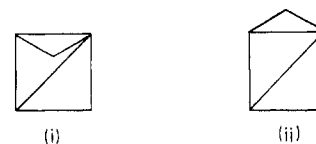


Figure 1.

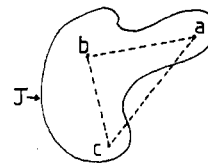


Figure 2.

edges, whereas the second projection has the infinite region bounded by five edges. This structure thus has two possible embeddings, corresponding to two distinct three-dimensional structures with different faces.

Topological transformation therefore involves a "rubber sheet geometry" approach, as described by Courant and Robbins,<sup>28</sup> whereby all projections of the same embedding can be transformed into one another by simple deformation, whereas transformation to a projection of a different embedding requires cutting or tearing followed by repair.

When there is no unique embedding, the problem of ring perception can become acute, especially for the definition of a particular ring set. This potential for topological ambiguity has not been appreciated in many of the ring perception techniques advocated.

Although a planar graph may have different embeddings, each with many different projections, the number of regions (finite and infinite) resulting from each projection of each embedding is the same and is given by Euler's polyhedral formula. Furthermore, in any simple, connected planar graph with  $f$  regions and more than two edges there must be at least one vertex with a connectivity of less than 6 and

$$e \geq \frac{3}{2}f \quad \text{and} \quad e \leq 3v - 6$$

must hold; otherwise, the graph is nonplanar.<sup>17</sup> The concept presented in this paper relies on the definition of regions of a planar graph and hence is not strictly applicable to nonplanar structures, although there are distinct parallels between the two.

Planar structures are covered by the theory of convex polytopes.<sup>29</sup> A **hull** is the boundary to a set of points, such as the Jordan curve,  $J$ , of Figure 2. A hull is **convex** if any two points within the hull can be joined by a line without crossing the boundary to the exterior. For instance, in Figure 2, points  $a$  and  $b$  and  $b$  and  $c$  can be joined legitimately, but points  $a$  and  $c$  cannot; thus, this is not a convex hull. A **polytope** is a hull containing a finite set of points. It is only the points which lie on the boundary that are of interest in the theory of polytopes; points on the interior or exterior are ignored.

A structure in which each vertex has a connectivity of at least  $d$  is said to be  **$d$ -connected**. A  $d$ -connected polytope is referred to as a  **$d$ -polytope**. The study of convex polygons (convex 2-polytopes) and polyhedra (convex 3-polytopes) has a very long history, being the subject of Greek geometry. Consideration of combinatorial aspects was started by Leonhard Euler, Jacob Steiner, and Arthur Cayley with respect to determining the numbers of nonisomorphic polyhedra. The regular polyhedra that Euler studied were convex polytopes, although a polyhedron does not have to be convex. Obviously a sphere is convex, and so the theory of convex polytopes can be applied legitimately in the following discussions on planar graphs.

Several general findings are of particular interest here, such as every block with at least three vertices is 2-connected, and

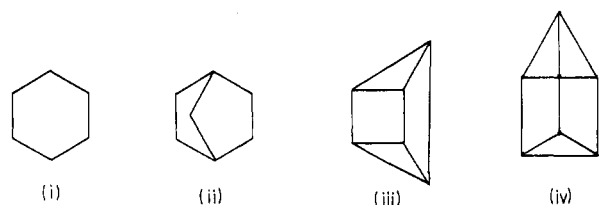


Figure 3.

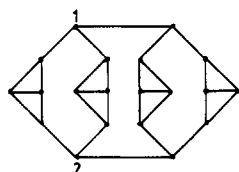


Figure 4.

Balinski's theorem<sup>30</sup> that every  $d$ -polytopal graph is  $d$ -connected. Grünbaum<sup>29</sup> uses the term "valent" for connectivity so that if there exist vertices of ring connectivity  $d + 1$  in the structure, then it is referred to as a  $d + 1$  valent  $d$ -polytope. To stress the  $d$ -valent nature over the general case, this is extended here to cover  $d$ -valent  $d$ -polytopes. For instance, in Figure 3, projection i is a 2-valent 2-polytope, projection ii is a 3-valent 2-polytope, projection iii is a 3-valent 3-polytope, and projection iv is a 4-valent 3-polytope. In general, projections i and ii are 2-polytopes, while projections iii and iv are 3-polytopes.

For 3-polytopal graphs, Steinitz<sup>31</sup> stated that "a graph  $G$  is polyhedral if and only if it is planar and 3-connected". This does not mean that a planar graph with every vertex with a ring connectivity of at least 3 is polyhedral. For instance, in Figure 4 removal of vertices 1 and 2 will result in an increase in the number of components (to three), even though each vertex has a ring connectivity of 3. Vertices 1 and 2 are thus cut vertices. Whitney studied this area and gives some of the fundamental results of graph theory:

"A graph  $G$  with at least  $d + 1$  nodes is  $d$ -connected if and only if every subgraph of  $G$ , obtained by omitting from  $G$  any  $d - 1$  or fewer nodes and the edges incident to them, is connected."

An important result that Whitney derived from Steinitz's theorem is that the faces of every 3-connected structure are uniquely determined by the planar graph. Hence, every polyhedron is uniquely embeddable on a sphere (or plane), and the regions of this embedding correspond to the faces of the polyhedron.

Note that the faces of a polyhedron are **coplanar**; i.e., all vertices of the face lie in the same plane. Chemical structures are inherently noncoplanar in that they are usually bent or twisted in some way. Polyhedra are a subset of the more general case. However, using a two-dimensional representation removes the problem and treats all faces as equivalent; i.e., if they form a region of any planar or spherical embedding, then they are equivalent, regardless of coplanarity. Where these faces are simple cycles, i.e., without Nachbarpunkte (see references 1 and 32), they will be referred to as **simple faces**.

#### EMBEDMENT CONSIDERATIONS

In mapping from three dimensions down to two dimensions most of the ring systems likely to be encountered in chemical information systems give a graph with a unique embedding. However, certain classes of complex ring system may give more than one embedding; i.e., the two-dimensional planar graph representation may correspond to more than one three-dimensional structure. Elk<sup>3</sup> discusses one such class and shows how different isomers of 4.3.2.1-paddlanes can lead to different

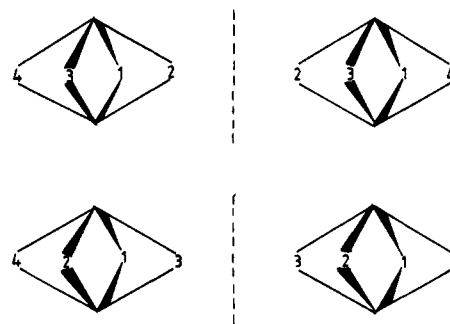


Figure 5.

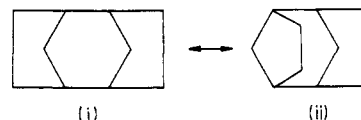


Figure 6.

planar embeddings (the 4, 3, 2, and 1 refer to the number of  $\text{CH}_2$  groups inserted into the rings of the paddlane). The four possible isomers result from two pairs of mirror images, as shown in Figure 5, with each pair having a different set of faces. In planar form there are thus two embeddings, each with two different projections; each embedding and projection will have the same number of regions, as given by the Euler formula, or finite regions, as given by the Cauchy formula, but the set of regions for each embedding will be different.

Original consideration of the problem of finding all alternative embedding regions came from the test structures given in Appendix A (Table III), referred to as the DBR database. While some, such as DBR-18 and DBR-23, have more than one embedding, most do not. The question arises as to what circumstances lead to ring systems having more than one embedding as planar (or spherical) projections.

Since any convex 3-polytope is uniquely embeddable, this is a possible upper limit to our problem—if a ring system is complex enough to be a 3-polytope, then we will only have one distinct embedding. The simplest possible convex 2-polytope is an isolated ring, i.e., a 2-valent 2-polytope. Embedded on a sphere or plane, it can have only one distinct embedding, giving two regions that have the same contour. This represents an initial lower bound for our problem. If vertices with a connectivity greater than 2 are introduced to our simplest 2-polytope, then, by the corollary to Euler's theorem, the number of vertices of odd connectivity must be even; i.e., there must be at least two of them. Similarly, for vertices of even connectivity greater than 2, if just one of them is introduced into a 2-valent 2-polytope (to give a spiro-fusion), then the graph is split into more than one block, with each block having a unique embedding. Thus, there must be at least two vertices of even ring connectivity greater than 2 for more than one embedding to be possible. The lower bound can now be more tightly defined as 2-polytopes with at least two vertices with a ring connectivity greater than 2. The upper bound is still a 3-polytope. A condition on the upper bound to consider is that if *all* vertices have a ring connectivity greater than 5, then the graph is nonplanar and the theory we are using is not appropriate.

It is now necessary to consider simple situations under which more than one embedding occurs. Figure 6 shows two embeddings of DBR-23. Embedment i has four regions giving two five-edged faces, a six-edged face, and an eight-edged face, while embedment ii also has four regions, but these give two five-edged and two seven-edged faces. The six- and eight-edged faces of embedment i and the two seven-edged faces of embedment ii are not interchangeable by simple deformation. As solid polyhedra, the change would require the pushing of

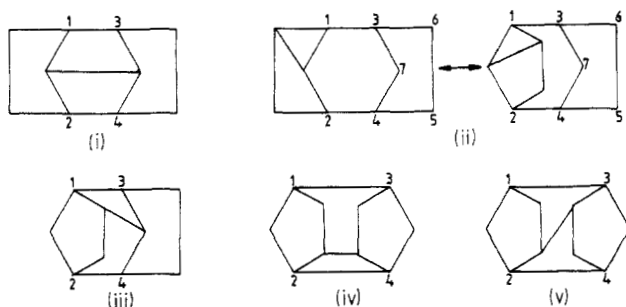


Figure 7.

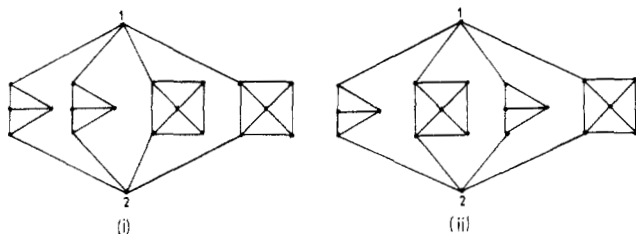


Figure 8.

one face through another; in terms of the models it requires the flipping of one *wing* over, or under, another. However, these structures are isomorphic, since incidence is preserved.

Now consider introducing a new edge to DBR-23 to join any pair of vertices with a connectivity of 2. Figure 7 shows the five possible ways of doing this. Notice that only structure ii has more than one embedding; all the rest have unique embeddings. Could it be that the presence of wings with vertices with a connectivity of 2 enables more than one embedding? In other words, is the property confined to structures with vertices with a ring connectivity of 2?

Consider the structure in Figure 8; all the vertices have a ring connectivity of more than 2, and yet structures i and ii are two different embeddings. Hence, the problem is not solely attributable to ring connectivity. Notice, however, that this structure is not a 3-polytope. Vertices 1 and 2 are cut vertices (articulation points); their removal, plus their incident edges, increases the number of components to four. *It is the presence of the pairs of cut vertices that allows parts of the structure to flip into a different embedding.* If there are no such pairs of cut vertices, then the structure is either a 2-valent 2-polytope or a 3-polytope and is thus uniquely embeddable. The presence of these pairs of cut vertices is thus a necessary condition for the occurrence of more than one embedding. However, as shown by structures i, iii, iv, and v of Figure 7, it is not a sufficient condition. Vertices 1 and 2, 3 and 4 are pairs of cut vertices (1 and 3, 1 and 4, etc. are not), but they do not allow another embedding. In Figure 7ii, 1 and 2, 3 and 4 do give rise to another distinct embedding. Note that in this structure these vertices have a connectivity of 3, with the pair 3 and 4 having the two unlinked paths, (3,7,4) and (3,6,5,4), joining them. In all the other structures there is only one unlinked path between vertices 3 and 4. For all structures, including structure ii, vertices 1 and 2 have at most one unlinked path between them.

It is suggested here that *it is the presence of more than one unlinked path between pairs of cut vertices that is sufficient to give more than one embedding.* An unlinked path (ULP) between a cut-vertex pair (CVP) is one that is not linked to any other edges incident to that CVP; otherwise, it is a linked path (LP). In other words, when tracing a path from one edge incident to one cut vertex, it is not possible to form a cycle containing that cut vertex without first having passed through the other cut vertex. *It can be seen immediately that unlinked paths occur from areas displaying 2-polytopal characteristics, while linked paths occur from areas displaying 3-polytopal*

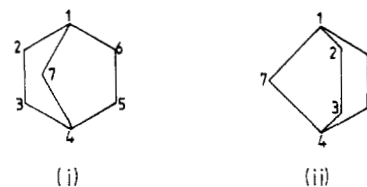


Figure 9.

*characteristics.* A further distinction can be made with respect to the simple cycles (i.e., regions of a planar graph) generated by the interaction between ULPs, ULPs, and LPs, and LPs. (There also exist simple cycles that cannot be generated by such interactions—these are the topic of the next section.) The simple cycles generated by these interactions can be divided into two categories:

- Those formed by ULPs and/or LPs incident with the same CVP will be referred to as **intra-CVP** combinations.
- Those that also include paths between, and leading to, other CVPs will be referred to as **inter-CVP** combinations.

Consider the structures given in Appendix A (Table III) to verify these suggestions; only norbornane (DBR-1) seems to be a problem. From Figure 9i it can be seen that 1 and 4 is a cut vertex pair and paths (1,2,3,4), (1,7,4), and (1,6,5,4) are unlinked. Therefore, by the above suggestions, norbornane should have more than one embedding. This is shown in Figure 9ii, but the regions defined are the same as in structure i. That structure ii is another distinct embedding can be shown to be true if one considers that Euler's formula predicts two faces for an isolated ring, a top side and a bottom side, but they have the same contour and hence give rise to only one distinct cycle. The two embeddings of norbornane can be thought of in the same way. Imagine norbornane with vertex 7 pushed down through the plane; one is left looking at the top side of the six-edged face. Embed this on a sphere and it can be rotated to give the bottom sides of two five-edged faces. Now pull vertex 7 above the plane; one is looking at the top side of the two five-edged faces. Embed on a sphere and rotation reveals the bottom side of the six-edged face. Hence, there are two distinct embeddings that look at different sides of a face. The sides of the faces may be different, but the contours are the same, and hence the ring sets are identical. The case of only one cut-vertex pair, with a ring connectivity of 3 and three ULPs connecting them, represents the lower bound for the occurrence of flipping embeddings!

The problem for ring perception is that from a two-dimensional connection table it is necessary to consider each possible three-dimensional realization to include all possible regions from all embeddings to find all possible simple faces. In discussing the problems associated with paddlanes, Elk<sup>3</sup> states

"For compounds that are intrinsically three-dimensional, the indiscriminate use of Schlegel diagrams can lead to major inconsistencies. In fact, the entire concept of projection is, at best, inappropriate for some compounds such as the paddlanes and buttaflanes."

This is not the case if all embeddings are considered. The development of the ESSR and the underlying theoretical basis of finding all possible alternative embedding regions refute Elk's statement.

#### CUT FACES

Including all possible alternative embedding regions is not always sufficient. There is an additional problem for substructure retrieval as shown by Figure 10. The file structure (i) is a heteroatom analogue of DBR-28. In the query

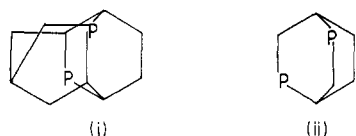


Figure 10.

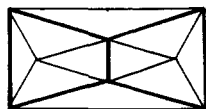


Figure 11.

structure (ii) all the six-edged simple cycles are symmetrically equivalent, and hence all should be included in the ring set. Unfortunately, in the file structure the six-edged simple cycle with two heteroatoms cannot be made into a region of any embedment and hence would not be included in the set of all simple faces. Failure is a result of this simple cycle in some way being "internal" to the file structure, so that it in effect cuts through the structure and has all its edges associated with more than one other simple cycle that can be regarded as "external" to the structure.

To extend the terminology already used in graph theory, any simple cycle that can be drawn as a region of some embedment of a structure will be called a **simple face**, as given earlier. Any simple cycle that cannot be drawn as a region of some embedment will be called a **cut face**. The simple faces correspond to the external simple cycles mentioned above and are the simple cycles that can be generated through the interaction of ULPs and LPs mentioned in the previous section. The cut faces correspond to the internal simple cycles mentioned above and are simple cycles that cannot be generated through the interaction of ULPs and LPs.

The infinite region of the maximal Schlegel projection also causes problems for substructure search. If it is larger than all the adjacent finite regions, then it should not be included in the query ring set since it will not necessarily occur in the ring set of the file structure, whereas all finite regions will. Hence, *if there is a unique maximal infinite region, then it should not be included in the query ring set*. Likewise, any cut faces that are symmetrically equivalent to such a maximal infinite region should be excluded from the ring set if they are similarly larger than all adjacent finite regions. For instance, the structure in Figure 11 has a four-edged maximal infinite region with two symmetrically equivalent cut faces (drawn in bold). All finite regions are three-edged, and so all the four-edged faces need to be excluded from the ring set if this structure is used as a substructure query.

The set of all simple cycles is thus divided into simple faces, which can form a region of some planar embedment, and cut faces, which cannot. Certain cut faces need to be included as members of a ring set to enable consistent substructure matching, while others need to be explicitly excluded. If all cut faces are included, then obviously the result is the set of all simple cycles! It is necessary therefore to investigate whether it is possible to define those cut faces that should be included and those that should not.

Cut faces can be divided into three classes: Class I contains those in which the cut face is smaller than at least one of its adjacent simple faces; i.e., it is the smallest ring for at least one of its edges. For instance, in Figure 12 the four-edged cut face (drawn in bold) is smaller than any of the adjacent regions and is the smallest ring in the structure. In these cases, although the cut face cannot become a region of the planar projection, it will occur in all SSSRs and hence in the set of  $\mathcal{H}$ -rings (see references 1 and 33). Class II contains those in which the cut face is the same size as at least one of its adjacent simple faces, but never smaller (i.e., it is a symmetrically

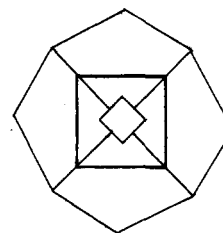


Figure 12.



Figure 13.

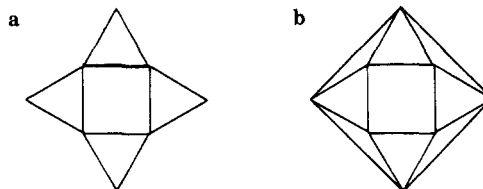


Figure 14.

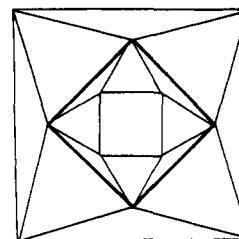


Figure 15.

equivalent ring for at least one of its edges), or is otherwise linearly independent of all its adjacent simple faces. Structures DBR-26, -28, -39, and -40 from Appendix A all contain such cut faces. Class III contains those in which all edges of the cut face are included in smaller simple faces and it is not linearly independent of them. For instance, in Figure 13 the six-edged cut face (drawn in bold) has all edges adjacent to the three four-edged simple faces, to which it is linearly dependent.

It can be seen from this that class I and class II cut faces are  $\mathcal{H}$ -rings and hence should be included in a ring set since they can be in at least one SSSR; they will be referred to as **primary cut faces**. Class III cut faces should not be included since they are not linearly independent of the simple faces associated with each edge; these will be referred to as **secondary cut faces**. Thus, primary cut faces are members of the set of  $\mathcal{H}$ -rings, while secondary cut faces are not.

It was mentioned in the review<sup>1</sup> that embedded ring systems present the problem that their SSSRs cannot be derived from any spanning tree. For instance, in Figure 14a the SSSR is (3,3,3,3,4), but no spanning tree exists whose chords generate this smallest set. If a spanning tree is produced whose chords generate the four three-edged simple faces, then the fifth chord must generate one of the larger non-simple-cycles and not the four-edged simple cycle. This is a problem for those techniques that rely on generation of a minimal spanning tree to generate an SSSR. For the structure in Figure 14b, by visual inspection the SSSR should be (3,3,3,3,3,3,3,4). There is a choice between two four-edged simple faces, neither of which can be derived from a spanning tree whose chords generate all the three-edged simple faces. The structure in Figure 15 continues the series. By visual inspection, the SSSR should be 16 three-edged simple faces and 1 four-edged. There are now

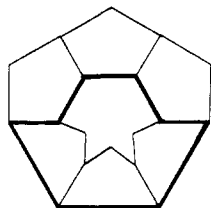


Figure 16.

three four-edged simple cycles to choose from; two are simple faces (the infinite and innermost finite regions), and one is a primary cut face (in bold). Each SSSR must contain one of them, and so the set of  $\mathcal{H}$ -rings contains all of them.

In Figure 16, the eight-edged primary cut face (drawn in bold) is smaller than the two nine-edged simple faces. For this structure the only SSSR is (5,5,5,5,5,8)—the nine-edged simple faces are thus not in the set of  $\mathcal{H}$ -rings.

Such potential problems with the SSSR and  $\mathcal{H}$ -ring ring sets are revealed by introducing the concept of simple faces and cut faces, and it is for this reason that the extended set of smallest rings is based on this concept to avoid such problems.

#### EXTENDED SET OF SMALLEST RINGS (ESSR)

The name ESSR is a result of the way in which the algorithm to find it has been developed; for historical reasons it has been kept in preference to the more correct "set of simple faces and primary cut faces". The ESSR contains

- all regions (finite and infinite) for every embedment, i.e., all simple faces, and
- all primary cut faces for every embedment.

The concept is applicable only to planar graphs due to its reliance on regions.

The relationship between the ESSR and the sets of all cycles, simple cycles, simple faces, cut faces, SSSRs,  $\mathcal{H}$ -rings, and  $\beta$ -rings (see references 1 and 34) is shown by the Venn diagram in Figure 17. A fundamental basis (see reference 1) can contain any cycle and so is an indeterminate subset of all cycles that cannot be conveniently represented. The essential set of essential rings (ESER) similarly cannot be conveniently represented due to the use of ideas of "synthetic importance" (see references 1, 35, and 36).

The proportions of the Venn diagram are not significant—it is the inclusion within and overlap between the different ring sets that is important. Similarly, the ring sets represent their respective general cases to indicate the possibility of a particular ring set including certain rings.

From the Venn diagram the following organization can be seen:

- The set of cut faces is divided into primary and secondary cut faces.
- The set of  $\mathcal{H}$ -rings includes all primary cut faces and a subset of the simple faces.
- An SSSR is a subset of the  $\mathcal{H}$ -rings that includes simple faces and possibly some primary cut faces.
- The set of  $\beta$ -rings includes a subset of all simple faces, all primary cut faces, and possibly some secondary cut faces.
- The ESSR contains all simple faces and all primary cut faces. Thus, it can be defined alternatively as the union of the set of simple faces and the set of  $\mathcal{H}$ -rings (as defined above).

The ESSR is the most appropriate ring set for file and query structures in a full-structure search and for file structures in a substructure search. For query structures in a substructure search only those members of the ESSR that are  $\mathcal{H}$ -rings can be used to ensure 100% recall, i.e., the set of  $\mathcal{H}$ -rings is the most appropriate ring set for substructure queries.

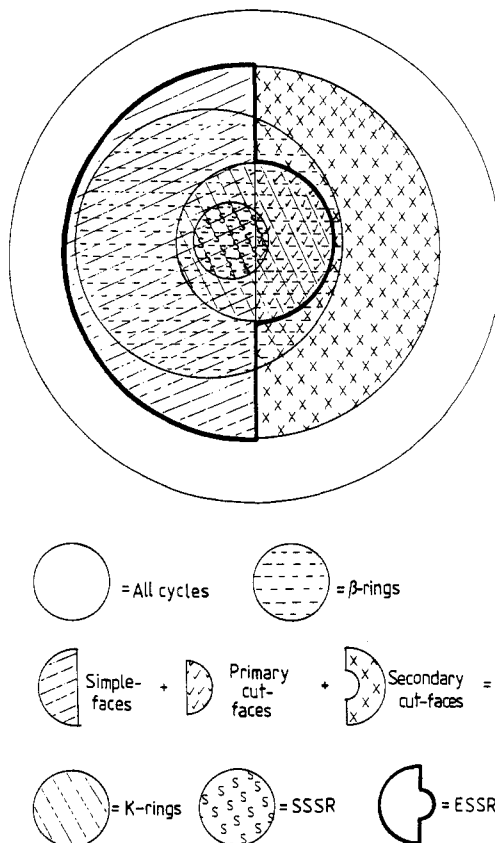


Figure 17.

Appendix A (Table III) gives a comparison between the rings included in the ESSR, an SSSR, and the sets of simple cycles,  $\mathcal{H}$ -rings, and  $\beta$ -rings for each of the DBR structures.

#### CAN THE ESSR CARDINALITY BE PREDICTED?

One of the most useful features of a fundamental basis, such as an SSSR, is that its cardinality can be calculated by the Cauchy formula, which simply requires the numbers of vertices and edges present. Thus, no knowledge of the rings or of their relationships is required. For simple structures, with a unique embedment and no primary cut faces, the only difference between the SSSR and the ESSR is the inclusion in the ESSR of the infinite region of the planar projection. For these structures the ESSR cardinality should be given by the Euler formula. However, for a single isolated ring the prediction would be for two rings, due to the two-faced nature of such rings, as mentioned earlier! In this special situation, the Cauchy formula is applicable.

The connection tables used on the Sheffield project benefit from a certain amount of preprocessing of the graph so that edges are denoted as being in a ring or chain. Thus, the information directly available from the connection table without any further ring perception or path tracing, includes

- the number of vertices,  $V$
- the number of edges,  $E$  (separated into ring or chain)
- the ring connectivity of each of the vertices
- the number of vertices of given ring connectivity
- the basis cardinality,  $\mu$  (Cauchy formula)
- the face cardinality (Euler formula), i.e.,  $\mu + 1$

The question is whether this information can be used to obtain a more accurate lower bound to the ESSR cardinality than the basis cardinality or even to give the exact ESSR cardinality. Table I lists these characteristics for the DBR database and for comparison includes the ESSR cardinality,  $\epsilon$ .

Although many different calculations have been attempted, it has not been possible to identify a formula that successfully

Table I. DBR Connection Table Characteristics

	<i>E</i>	<i>V</i>	ring connectivity					$\mu$	$\epsilon$		<i>E</i>	<i>V</i>	ring connectivity					$\mu$	$\epsilon$
			2	3	4	5	6						2	3	4	5	6		
DBR-1	8	7	5	2	-	-	-	2	3	DBR-21	17	11	-	10	1	-	-	7	8
DBR-2	9	8	6	2	-	-	-	2	3	DBR-22	21	13	2	6	5	-	-	9	10
DBR-3	8	5	-	4	1	-	-	4	5	DBR-23	12	10	6	4	-	-	-	3	6
DBR-4	15	13	9	4	-	-	-	3	4	DBR-24	12	10	8	2	-	-	-	3	6
DBR-5	12	8	-	8	-	-	-	5	6	DBR-25	47	34	22	8	4	-	-	9	12
DBR-6	9	6	-	6	-	-	-	4	5	DBR-26	19	14	4	10	-	-	-	6	7
DBR-7	13	11	7	4	-	-	-	3	5	DBR-27	11	9	5	4	-	-	-	3	4
DBR-8	12	10	6	4	-	-	-	3	4	DBR-28	15	12	6	6	-	-	-	4	6
DBR-9	13	11	7	4	-	-	-	3	4	DBR-29	24	20	16	-	4	-	-	5	5
DBR-10	18	12	-	12	-	-	-	7	8	DBR-30	28	22	18	-	4	-	-	5	15
DBR-11	30	24	12	12	-	-	-	7	8	DBR-31	24	20	18	-	-	-	2	5	15
DBR-12	24	18	6	12	-	-	-	7	8	DBR-32	22	19	17	-	-	2	-	4	10
DBR-13	16	12	4	8	-	-	-	5	5	DBR-33	16	13	11	-	-	2	-	4	10
DBR-14	12	8	4	4	-	-	-	5	5	DBR-34	14	12	10	-	2	-	-	3	6
DBR-15	16	12	8	4	-	-	-	5	5	DBR-35	14	12	10	-	2	-	-	3	5
DBR-16	13	12	10	2	-	-	-	2	3	DBR-36	12	6	-	-	6	-	-	7	8
DBR-17	14	12	8	4	-	-	-	3	4	DBR-37	24	19	10	8	1	-	-	6	8
DBR-18	10	8	6	2	-	-	-	3	6	DBR-38	23	18	10	6	2	-	-	6	8
DBR-19	13	10	6	4	-	-	-	4	7	DBR-39	22	17	8	8	1	-	-	6	9
DBR-20	16	11	1	10	-	-	-	6	7	DBR-40	18	14	6	8	-	-	-	5	7

predicts the ESSR cardinality or finds a better lower bound using these characteristics alone. More information is required, for which it is necessary to consider the underlying theory behind the number of rings in any given ESSR. The rest of this section studies the theory and develops a means of prediction for the number of simple faces in any ESSR to provide a better lower bound to the ESSR cardinality than the nullity.

First, it ought to be stated that the number of primary cut faces cannot be predicted. Although the total number of cut faces can be calculated, whether each cut face is primary or secondary is a feature of the ring sizes of the particular structure under consideration. It therefore requires the perception of all adjacent simple faces first, which, of course, removes the predictive element! Thus, the problem can be confined to prediction of the number of distinct regions resulting from all possible embeddings of a structure, i.e., prediction of the number of all possible simple faces—the simple face cardinality.

The first assumption is that the individual blocks of a graph are processed separately. For a structure such as the spiro-fused adamantanes in DBR-37 the blocks correspond to the two adamantanes. Processing each one individually is simpler than processing them together and is more efficient, as shown by the separate processing of blocks in Wipke and Dyott's algorithm.<sup>37</sup> As mentioned above, if all vertices in a block have a ring connectivity of 2, then it is a single ring (which adds 1 to the ESSR cardinality, not 2). In this special case there is no possibility of there being a cut face, and so the simple-face cardinality and ESSR cardinality are the same. Thus, there is no problem in correctly predicting the ESSR cardinality for 2-valent 2-polytopes.

If one vertex of even ring connectivity ( $>2$ ), such as a spiro-fusion, is added, then this vertex constitutes a cut vertex and the graph can be split into two separate blocks. If two or more such vertices are added to a single block, then, if it remains a single block, the ESSR cardinality becomes more difficult to predict due to the possibility of other embeddings and primary cut faces. Recalling that the number of vertices of odd connectivity must be even, if one adds two odd-connectivity vertices (or more) to a block, then once again the ESSR cardinality becomes difficult to predict.

*If a structure is a 2-valent 2-polytope, or can be reduced to such by separating the structure into its blocks, then the ESSR cardinality is given by the Cauchy formula.*

*If a structure is a 3-polytope, then it has a unique embedding and the simple-face cardi-*

*nality is predicted by the Euler formula.*

In between are the  $(2+n)$ -valent 2-polytopes for which the number of simple faces is more difficult to predict due to the possibility of more than one embedding. Hence, it is the *lower bound of the simple-face cardinality* that is predicted by the Euler formula for  $(2+n)$ -valent 2-polytopes. Additional rings can arise from the presence of alternative embeddings and/or primary cut faces, but there cannot be fewer rings, and so this also represents the *lower bound of the ESSR cardinality*. Similarly for 3-polytopes the lower bound of the ESSR cardinality is predicted by the Euler formula, but additional rings can arise only from the structure having primary cut faces.

From the information directly available, such as that given in Table I, it is not possible to predict the occurrence of additional embeddings or cut faces. However, it was shown earlier that more than one embedding results from the presence of pairs of cut vertices with more than one unlinked path between them. Let us assume that, by a certain amount of additional processing, it is possible to detect the pairs of cut vertices and find the unlinked paths between them. Given this information, it is possible to predict the total number of simple faces resulting from all distinct embeddings?

In the following discussion it is necessary to introduce a new concept, that of the **cut-vertex graph**. This is similar to the vertex graphs used by Lederberg and others<sup>5-8</sup> in that vertices with a ring connectivity of two are condensed (contracted) into an arc, i.e., the basic graph is produced in which vertices represent atoms with a ring connectivity of 3 or more. The difference is that for the cut-vertex graph concept introduced here an additional constraint is that the vertices represent cut vertices only.

*The arcs of a cut-vertex graph represent either unlinked paths (ULPs) or boundaries to regions that contain linked paths (LPRs).*

An LPR is formed when at least two arcs incident to a cut-vertex pair (CVP) are linked to each other by edges not incident to the CVP. If there are more than two dependent arcs in such a region, then any arcs contained within the outer arcs can be ignored since they can never form simple faces with any ULPs; i.e., they can never combine with a ULP to form a region of any embedding. This can be visualized more easily by attempting to transform the graph so that an inner linked-path arc combined with a ULP becomes the infinite region. In Figure 18i, arc b through the LPR is not free to move to the other side of arcs a or c and hence can never be part of an infinite region. Note also that the combination of both boundary arcs, ac, is similarly unable to form the infinite

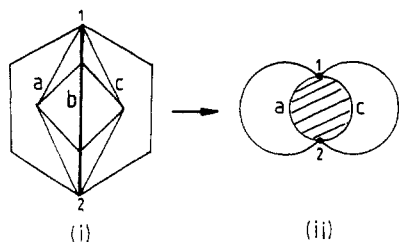


Figure 18.

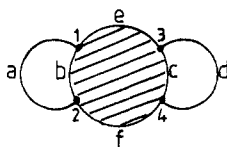


Figure 19.

region. Thus, LPRs can be represented just by their boundary arcs, as shown in Figure 18ii.

By definition, an LPR contains no pairs of cut vertices. Hence, an isolated LPR is a 3-polytope, has a unique embedding, and, since the boundary defines the infinite region, contains the number of simple faces predicted exactly by the Euler formula. If, however, an LPR has one or more incident CVPs, then the boundary, as shown in Figure 18, cannot form the infinite region. It can only contribute to the number of simple faces by combination with the adjacent ULPs, and so the number of simple faces in any LPR having incident CVPs is given exactly by the Cauchy formula.

*An LPR with no incident CVPs has a simple-face cardinality calculated by the Euler formula; otherwise, its simple-face cardinality is given by the Cauchy formula (i.e., the nullity).*

In cut-vertex graphs, LPRs are shown as shaded areas bounded by as many arcs as there are incident cut vertices. None of the inner connections of the LPR are shown. All ULP and LPR arcs connecting CVPs are labeled alphabetically. ULP arcs by definition must join some CVP, but it is possible to have LPR arcs that do not. For instance, in Figure 19 the LPR has two incident CVPs, (1,2) and (3,4). The pairs (1,3) and (2,4) are not CVPs, although their arcs are labeled because they can contribute to a region.

If the fusion points between two or more simple cycles are Nachbarpunkte, then it is not possible for a further simple cycle to include vertices from two or more of these simple cycles and the Nachbarpunkte fusion points. Similarly in a cut-vertex graph it is possible to have CVPs that are Nachbarpunkte. These are shown as straight lines connecting cut vertices rather than as arcs. These CVPs cannot contribute to alternative embedding regions of the cut-vertex graph, and so the graph can be simplified by their removal. This can be achieved by contracting all Nachbarpunkte CVPs to a single vertex, which allows simplification to further blocks. The concept of contraction of edges between vertices of a "vertex graph", i.e., a basic graph, has already been outlined by Carhart et al.<sup>6</sup> (who use the term "collapse" instead of contraction). If contraction is applied to cut-vertex graphs, the results can include self-loops, multiple edges, and contraction to a single vertex. An alternative that avoids these complications is to concentrate on the nature of the edge between the Nachbarpunkte cut vertices. This edge is a cut edge, the cutting of which enables the separation of the cut-vertex graph into further blocks. Separation of these CVPs thus involves cutting the graph along the length of the intervening edge and replacing the CVPs by an arc in both the blocks formed by the separation. For the DBR test structures most of the resultant blocks are thereby reduced to 2-valent 2-polytopes, each

Table II. Examples of Nachbarpunkte Cut-Vertex Pairs

cut-vertex graph	Nachbarpunkte contracted	Nachbarpunkte separated
1 (DBR-17)		
2 (DBR-29)		
(DBR-14)		
3 (DBR-22)		

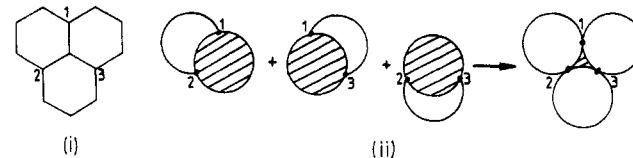


Figure 20.

of which can contribute no more than one simple cycle to the ESSR.

Nachbarpunkte CVPs can occur in three situations, as outlined in Table II. This table contrasts the results obtained from contraction and separation for certain of the DBR structures (DBR-17, -29, -14, and -22, respectively). All four examples result in the occurrence of self-loops under contraction, while the second and third examples contract to multiple edges and a single vertex. Separation of the cut-vertex graph yields much simpler blocks in all cases. Note that in the fourth example separation yields a 3-polytopal block for which the number of simple faces is given by the Cauchy formula, not the Euler formula, since the 3-polytopal status has been achieved by separation of the incident CVPs.

A single edge can also connect two cut vertices of different CVPs. These edges are not between Nachbarpunkte CVPs and so are not cut edges that can be used to separate the cut-vertex graph into simpler blocks. Such non-CVP Nachbarpunkte can occur incident to areas displaying either 2-valent 2-polytopal or 3-polytopal characteristics. Unlike the CVP Nachbarpunkte, it is not necessary to represent the non-CVP Nachbarpunkte by straight edges rather than arcs, since, as there is only ever one path directly linking them, they cannot contribute to alternative embedding regions other than those formed by the interaction of the Nachbarpunkte CVPs.

To differentiate LPRs it is necessary to consider all CVPs. What may be thought of as an LPR with respect to one CVP



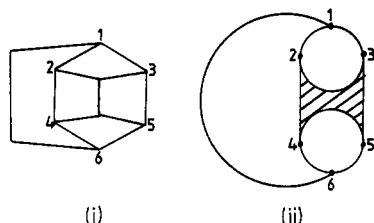


Figure 21.

may not necessarily be 3-polytopal, thus making it possible to define further CVPs and ULPs. For instance, DBR-4, Figure 20i, is a composite of three CVPs. Figure 20ii shows the viewpoints from the individual CVPs (1,2), (1,3), and (2,3) leading to the overlaid situation for all three. As can be seen, when all three CVPs are considered, the LPR effectively reduces to the central vertex and incident edges of (i), with a nullity of 0. The CVPs have common vertices, but the ULPs do not overlap to give common parts of an arc. Hence, the ULPs from these CVPs can be said to be *independent* of one another.

In contrast, the structure in Figure 21i has a cut-vertex graph (ii) comprising an LPR with two incident CVPs and two non-CVP single-vertex boundary edges, four non-CVP single-edged ULPs, and one ULP arc. In this case the ULPs between the CVPs (2,3) and (4,5) are interrupted by the CVP (1,6), and hence these ULPs can be said to be *dependent*.

The simplification of a graph to its cut-vertex form is ideally suited to revealing the combinatorial relationship between ULPs, LPRs, and the simple-face cardinality. This relationship has been revealed by consideration of 146 cut-vertex graphs comprising 20 series of increasing complexity, ranging from simple combinations of ULPs around a single CVP to mixtures of independent and dependent ULPs and LPRs with large numbers of CVPs. Although not exhaustive, these series are sufficient to show that the underlying combinatorial relationship between the arcs and regions of the cut-vertex graph generates all the alternative embedment regions corresponding to the equivalent simple faces of the associated ESSR. A sample cut-vertex graph from each of the 20 series is given in Appendix B (Table IV).

The general conclusions drawn from consideration of the series of cut-vertex graphs are as follows:

- If a cut-vertex graph is a 2-valent 2-polytope, or a series of 2-valent 2-polytopes, then each such block contributes 1 to the number of regions (simple faces) in the graph. The same is true if the blocks are a result of separation of Nachbarpunkte CVPs.
- If a cut-vertex graph is a 3-polytope, or a series of 3-polytopes, then each contributes a number of regions calculated by the Euler formula for each such block. If a 3-polytopal block is a result of the separation of incident Nachbarpunkte CVPs, then it will contribute a number of regions calculated by the Cauchy formula.
- For blocks that are not just 2-valent 2-polytopal or 3-polytopal, the number of alternative embedment regions corresponds to the sum of the intra-CVP combinations and the inter-CVP combinations.
- The number of intra-CVP regions formed by incident ULPs is calculated as the number of combinations of two of these ULPs. When an LPR is incident to a CVP, if this is the only CVP incident to that LPR, then both LPR boundary arcs can contribute to combinations with any incident ULPs, but the combination of both LPR boundary arcs is not valid since it forms a cut face. This will be referred to as a **primary** LPR. CVG-2 and -3 illustrate the subtraction of one from the intra-CVP combinations of each

primary LPR, to which both boundary arcs contribute. Where there is more than one incident CVP to an LPR, then each boundary arc contributes only to combinations with the ULPs incident to the associated CVP. This will be referred to as a **secondary** LPR. A secondary LPR effectively contributes only one arc to all intra-CVP combinations. CVG-7 illustrates the subtraction of one arc from the factorial calculation for each secondary LPR; i.e., only one boundary arc is included.

- The number of inter-CVP regions is calculated by summing the products of the number of ULPs or primary LPR arcs between each cut vertex in each inter-CVP combination.

Where CVPs are contained within other pairs, as shown by CVG-13 and -14, they can be regarded as nested subproblems. In this way added complexity can be dealt with by taking the sum and product of less complex subgraphs of the cut-vertex graph.

The general formula for the number of regions of a cut-vertex graph that corresponds to the number of simple faces in the full graph is

- all valid intra-CVP combinations, plus
- all valid inter-CVP combinations, plus
- all primary and secondary LPR nullities, plus
- 1 for each 2-valent 2-polytopal block, plus
- the nullity of each 3-polytopal block resulting from Nachbarpunkte CVP separation, plus
- the Euler number from each 3-polytope

where "valid" is a function of any LPRs that may be present and is determined by their primary or secondary nature.

## SUMMARY

This paper has investigated the topological principles necessary to define a new ring set, the extended set of smallest rings, that overcomes the ambiguities associated with other ring sets, as revealed by the previous review of ring perception techniques.

The differentiation of the set of all simple cycles into simple faces and cut faces (primary and secondary) enables a direct generalized comparison between the ESSR and most of the other ring sets in a way that highlights the areas of difference clearly and understandably.

The importance of considering alternative embedment regions and cut faces has been discussed in detail. The theory of convex polytopes is particularly useful in explaining the conditions necessary for the occurrence of alternative embedment regions. The idea of cut-vertex pairs with incident unlinked paths and linked path regions has led to the cut-vertex graph concept. The cut-vertex graph reveals the combinatorial relationship between the full graph and all its possible regions (simple faces) and conceptually provides a simple formalism for the description of the whole range of possible ring systems.

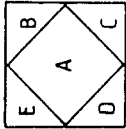
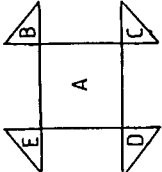
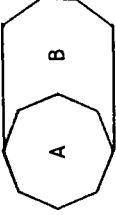
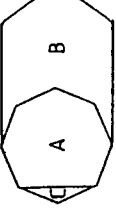
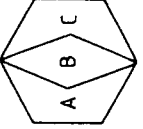
An appendix is given in which the ESSR is compared with various other ring sets for 40 specific structures mostly gleaned from the literature as problem structures for other ring perception programs. A second appendix presents 20 representative cut-vertex graphs from the series used to explain the combinatorial relationship between the cut-vertex graph and all possible regions.

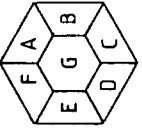
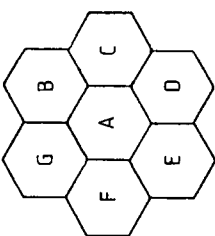
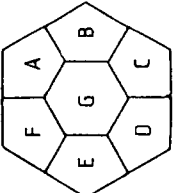
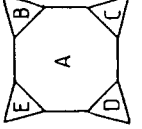
The algorithms developed to find the ESSR within specific and structurally explicit generic structures and to represent the rings as bit vectors in the correct logical relationships are given in subsequent papers.<sup>38,39</sup>

## ACKNOWLEDGMENT

This research was made possible through funding provided by IDC (Internationale Dokumentationsgesellschaft für



1		2		3			4			5		
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note
DBR-14 	$\mu = 5$ B,C,D,E A A+B+C+D+E											
DBR-15 	$\mu = 5$											
DBR-16 	$\mu = 2$ A B A+B											
DBR-17 	$\mu = 3$ C A B A+B A+B+C											
DBR-18 	$\mu = 3$ B A,C A+B, B+C A+B+C											
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note

1		2		3			4			5		
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note
DBR-10 	$\mu = 7$ A, ..., F G A+B+...+G A+B etc. A+B+G etc. A+B+C+G etc.											
DBR-11 	$\mu = 7$ A B, ..., G A+B+...+G A+B+C etc. A+B+C+D etc.											
DBR-12 	$\mu = 7$ A, ..., F G A+B+...+G A+B+G etc. A+B+C+G etc.											
DBR-13 	$\mu = 5$ B,C,D,E A A+B+C+D+E											
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note



1	2	3					4					5		
		G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$		Note	
DBR-29	$\mu = 5$ Structure diagram 	Cycle type												
		C	4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		B,D	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (F)
		A+B, D+E	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (Y)
		A+B+C, B+C, C+D+E, C+D	8	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (N)
		A,E	10	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (N)
		A+B+C+D, B+C+D+E, A+B+C+D+E, B+C+D+E	12	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (N)
DBR-30	$\mu = 5$ Structure diagram 													
		C	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		B,D	7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (F)
		A+B, D+E	7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (Y)
		A+B+C, B+C, C+D+E, C+D	9	✓	✓	x	x	✓	✓	✓	✓	✓	✓	4 0 0 4 (R)
		A,E	10	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 2 (F)
		A+B+C+D, A+B+C+D+E, B+C+D, B+C+D+E	12	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 4 (I)
DBR-31	$\mu = 5$ Structure diagram 													
		C	4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		B,D	7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (F)
		A+B, D+E	7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (Y)
		A+B+C, B+C, C+D+E, C+D	7	✓	✓	x	x	✓	✓	✓	✓	✓	✓	4 0 0 4 (Y)
		A,E	10	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 2 (F)
		A+B+C+D, A+B+C+D+E, B+C+D, B+C+D+E	10	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 4 (I)
		Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note	

1	2	3					4					5		
		G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$		Note	
DBR-26	$\mu = 6$ Structure diagram 	Cycle type												
		A,D	4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (F)
		C,E	5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	2 2 2 2 (F)
		B	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		A+B+...+F	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (O)
		B+C+D+E	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 0 1 1 (R)
		B+A+F	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 0 1 1 (P)
		A+B+C+F, A+B+E+F	7	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (S)
		A+B+C+D+F, A+B+D+E+F	7	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (S)
		F	8	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (N)
		B+C+D, B+D+E	9	x	x	x	x	x	x	x	x	x	x	0 0 0 0 (S)
DBR-27	$\mu = 3$ Structure diagram 													
		A	4	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		B	5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		C	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		A+B+C	7	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 1 (M)
		B+C	7	✓	✓	x	x	✓	✓	✓	✓	✓	✓	1 1 0 0 (S)
		A+B	7	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (N)
		A+C	8	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (N)
DBR-28	$\mu = 4$ Structure diagram 													
		D	5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 1 1 1 (F)
		A,B,C	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	3 3 3 3 (F)
		B+C	6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1 0 1 1 (P)
		A+B+C+D	7	✓	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 1 (M)
		A+D	7	✓	✓	x	x	✓	✓	✓	✓	✓	✓	1 0 0 0 (S)
		A+B+C	8	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (S)
		A+C+D	9	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (S)
		C+D, B+C+D	9	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (N)
		A+B, A+C	10	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (N)
		A+B+D	11	x	x	x	x	✓	✓	✓	✓	✓	✓	0 0 0 0 (N)
		Cycle type	G	e	$\beta$	S	K	$\epsilon$	$\beta$	S	K	$\epsilon$	Note	

Table III (Continued)

1		2		3			4			5		
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\mathcal{E}$	$\beta$	S	K	$\mathcal{E}$	Note
DBR-32 	$\mu = 4$ B,C A+B, C+D A+B+C, B+C, A+B+C+D, B+C+D A,D	7 7 10 10	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	2 2 4 2	2 2 0 0	2 2 0 2	2 2 0 2	(F) (Y) (I) (F)
DBR-33 	$\mu = 4$ B,C A+B, C+D A,D A+B+C, B+C+D A+B+C+D	4 4 7 7 7 10	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\times$	2 2 2 2 2 0	2 2 1 2 2 0	2 2 2 2 2 0	2 2 2 2 2 0	(F) (Y) (A) (Y) (M)
DBR-34 	$\mu = 3$ B A,C A+B, B+C A+B+C	4 7 7 10	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$	1 2 2 0	1 2 2 0	1 2 2 0	1 2 2 0	(F) (F) (Y) (M)
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\mathcal{E}$	$\beta$	S	K	$\mathcal{E}$	Note
DBR-35 	$\mu = 3$ B C A+B, B+C A A+B+C	5 6 7 8 9	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$	$\checkmark$ $\checkmark$ $\times$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	1 1 2 0 0	1 1 2 0 0	1 1 2 0 0	1 1 2 0 0	(F) (F) (A) (F) (M)
DBR-36 	$\mu = 7$ A, ..., G A+B+...+G A+B+C+F, B+C+D+E A+B+D+G A+B etc. B+E etc.	3 3 4 4 5	$\checkmark$ $\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\checkmark$ $\times$ $\times$	7 7 7 3 0	7 7 7 0 0	7 7 7 3 0	7 7 7 0 0	(F) (I) (S) (N) (D)
DBR-37 	$\mu = 6$ A, ..., F A+B+C, D+E+F A+B etc. A+B+C+D etc. A+B+...+F	6 6 8 12	$\checkmark$ $\checkmark$ $\checkmark$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	$\checkmark$ $\checkmark$ $\times$ $\times$	6 2 6 0	6 2 6 0	6 2 6 0	6 2 6 0	(F) (I) (S) (D)
Structure diagram	Cycle type	G	e	$\beta$	S	K	$\mathcal{E}$	$\beta$	S	K	$\mathcal{E}$	Note



Table IV.

Cut-vertex graph	Arc combinations	$\mathcal{F}$	Combinatorial equation
<p>CVG-1</p>	<p><i>ab bc cd de</i>  <i>ac bd ce df</i>  <i>ad be cf</i>  <i>ae bf</i>  <i>af ef</i></p>	<p>15</p>	<p><math>\frac{6!}{2 \cdot 4!} = 15</math></p>
<p>CVG-2</p>	<p><i>(ab) bc cd de</i>  <i>ac bd ce df</i>  <i>ad be cf</i>  <i>ae bf</i>  <i>af ef</i></p>	<p><math>14 + \mu^{ab}</math></p>	<p><math>\frac{6!}{2 \cdot 4!} - 1 = 14</math></p>
<p>CVG-3</p>	<p><i>(ab) bc cd (de)</i>  <i>ac bd ce df</i>  <i>ad be cf</i>  <i>ae bf</i>  <i>af ef</i></p>	<p><math>13 + \mu^{ab} + \mu^{de}</math></p>	<p><math>\frac{6!}{2 \cdot 4!} - 2 = 13</math></p>
<p>CVG-4</p>	<p><i>ab bc de ef</i>  <i>ac df</i>  <i>agdh bgdh cgdh</i>  <i>ageh bgeh cgeh</i>  <i>agfh bgfh cgfh</i></p>	<p>15</p>	<p><math>\frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + [(3 \cdot 1 \cdot 3 \cdot 1)] = 15</math></p>
<p>CVG-5</p>	<p><i>ab cd ef</i>  <i>agci bgci chej dhej</i>  <i>agdi bgdi chfj dhfj</i>  <i>agheji bgheji</i>  <i>aghfji bghfji</i></p>	<p>15</p>	<p><math>\frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} + 0 + 0 + 0 + 0 + \frac{2!}{2 \cdot 0!} + [(2 \cdot 1 \cdot 2 \cdot 1) + (2 \cdot 1 \cdot 2 \cdot 1) + (2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1)] = 15</math></p>
<p>CVG-6</p>	<p><i>ab ef chdk</i>  <i>agcj bgcj</i>  <i>dief difl</i>  <i>aghdkj bghdkj</i>  <i>chielk chiflk</i>  <i>aghielkj bghielkj</i>  <i>aghiflkj bghiflkj</i></p>	<p>15</p>	<p><math>\frac{2!}{2 \cdot 0!} + 0 + 0 + 0 + \frac{2!}{2 \cdot 0!} + 0 + 0 + 0 + [(2 \cdot 1 \cdot 1 \cdot 1) + (1 \cdot 1 \cdot 1 \cdot 1) + (1 \cdot 1 \cdot 1 \cdot 2) + (2 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1) + (1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1) + (2 \cdot 1 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1)] = 15</math></p>



Table IV (Continued)

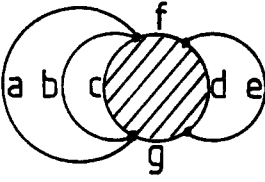
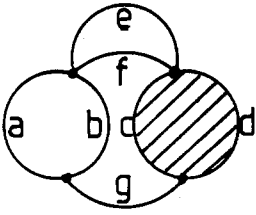
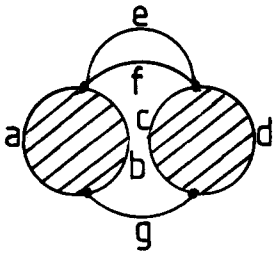
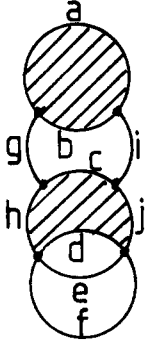
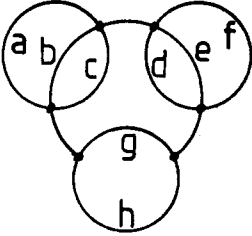
Cut-vertex graph	Arc combinations	$\mathcal{F}$	Combinatorial equation
<p>CVG-7</p> 	<p> <math>ab \quad bc \quad de</math>  <math>ac</math>  <math>(afd_g) \quad (bfd_g) \quad (cfd_g)</math>  <math>afeg \quad bfeg \quad cfeg</math> </p>	<p><math>6 + \mu^{cfd_g}</math></p>	<p><math>\frac{3!}{2 \cdot 1!} + \frac{2!}{2 \cdot 0!} + [(2.1.1.1)] = 6</math></p>
<p>CVG-8</p> 	<p> <math>ab \quad (cd) \quad ef</math>  <math>aecg \quad becg</math>  <math>aedg \quad bedg</math>  <math>afcg \quad bfcg</math>  <math>afdg \quad bfdg</math> </p>	<p><math>10 + \mu^{cd}</math></p>	<p><math>\frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} + \left(\frac{2!}{2 \cdot 0!} - 1\right) + 0 + [(2.2.2.1)] = 10</math></p>
<p>CVG-9</p> 	<p> <math>(ab) \quad (cd) \quad ef</math>  <math>aecg \quad becg</math>  <math>aedg \quad bedg</math>  <math>afcg \quad bfcg</math>  <math>afdg \quad bfdg</math> </p>	<p><math>9 + \mu^{ab} + \mu^{cd}</math></p>	<p><math>\left(\frac{2!}{2 \cdot 0!} - 1\right) + \frac{2!}{2 \cdot 0!} + \left(\frac{2!}{2 \cdot 0!} - 1\right) + 0 + [(2.2.2.1)] = 9</math></p>
<p>CVG-10</p> 	<p> <math>(ab) \quad de \quad ef</math>  <math>agci \quad bgci</math>  <math>(chdj) \quad (chej) \quad (chfj)</math>  <math>(aghdji) \quad agheji \quad aghfgi</math>  <math>(bghdji) \quad bgheji \quad bghfgi</math> </p>	<p><math>9 + \mu^{ab} + \mu^{chdj}</math></p>	<p><math>\left(\frac{2!}{2 \cdot 0!} - 1\right) + 0 + 0 + 0 + \frac{3!}{2 \cdot 1!} + [(2.1.1.1)] + (0.1.2.1) + (2.1.1.2.1.1)] = 9</math></p>
<p>CVG-11</p> 	<p> <math>ab \quad bc</math>  <math>ac \quad de \quad ef</math>  <math>df \quad gh</math>  <math>adg \quad bdg \quad cdg</math>  <math>adh \quad bdh \quad cdh</math>  <math>aeg \quad beg \quad ceg</math>  <math>afh \quad beh \quad ceh</math>  <math>afg \quad bfg \quad cfg</math>  <math>afh \quad bfh \quad cfh</math> </p>	<p>25</p>	<p><math>\frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + \frac{2!}{2 \cdot 0!} + [(3.3.2)] = 25</math></p>

Table IV (Continued)

Cut-vertex graph	Arc combinations	$\mathcal{F}$	Combinatorial equation
CVG-12 	$ab$ $bc$ $ef$ $ac$ $de$ $gh$ $(adg)$ $(bdg)$ $(cdg)$ $(adh)$ $(bdh)$ $(cdh)$ $(aeg)$ $(beg)$ $(ceg)$ $(afh)$ $(bfh)$ $(cfh)$	11	$\frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + \frac{2!}{2 \cdot 0!} + [(2.2.1)] = 11$
CVG-13 	$ab$ $ac$ $bc$ $de$ $ef$ $hi$ $df$ $gh$ $gi$ $adg$ $bdg$ $cdg$ $adh$ $bdh$ $cdh$ $adi$ $bdi$ $cdi$ $aeg$ $beg$ $ceg$ $afh$ $bfh$ $cfh$ $afi$ $bfi$ $cfi$	36	$\frac{2!}{2 \cdot 0!} + \left( \frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + 0 \right) + [(2.1) + (2.3.3) + (1.3.3)] = 36$
CVG-14 	$ab$ $ac$ $(bc)$ $de$ $ef$ $hi$ $df$ $gh$ $gi$ $(adg)$ $(bdg)$ $(cdg)$ $(adh)$ $(bdh)$ $(cdh)$ $(adi)$ $(bdi)$ $(cdi)$ $(aeg)$ $(beg)$ $(ceg)$ $(afh)$ $(bfh)$ $(cfh)$ $(afi)$ $(bfi)$ $(cfi)$	12 + $\mu^{bdg}$	$\frac{2!}{2 \cdot 0!} + \left( \frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + 0 \right) + [(1.1) + (1.0.0) + (1.2.2)] = 12$
CVG-15 	$ab$ $bc$ $de$ $ac$ $ef$ $df$ $gh$ $hi$ $ij$ $gi$ $hj$ $gj$ $ajd$ $bjd$ $cjd$ $aje$ $bje$ $cje$ $ajf$ $bjf$ $cjf$ $ai.$ $bi.$ $ci.$ $ah.$ $bh.$ $ch.$ $ag.$ $bg.$ $cg.$	48	$\frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + \left( \frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} \right) + [(3.3.2) + (3.3.2) + (2.2)] = 48$
CVG-16 	$ab$ $bc$ $de$ $ac$ $ef$ $df$ $gh$ $hi$ $(ij)$ $gi$ $hj$ $gj$ $ajd$ $bjd$ $(cjd)$ $aje$ $bje$ $(cje)$ $(ajf)$ $(bjf)$ $(cjf)$ $ai.$ $bi.$ $(ci.)$ $ah.$ $bh.$ $(ch.)$ $(ag.)$ $(bg.)$ $(cg.)$	23 + $\mu^{cgf}$ + $\mu^{ij}$	$\frac{3!}{2 \cdot 1!} + \frac{3!}{2 \cdot 1!} + \left( \frac{2!}{2 \cdot 0!} + \left( \frac{2!}{2 \cdot 0!} - 1 \right) \right) + [(2.2.1) + (2.2.2) + (2.2)] = 23$

Table IV (Continued)

Cut-vertex graph	Arc combinations	$\mathcal{F}$	Combinatorial equation
CVG-17 	$ab \quad ef \quad gh$ $cde \quad cdf$ $ijg \quad ijh$ $acki \quad bcki$ $adlj \quad bdlj$ $ekgl \quad fkgi$ $ekhl \quad fkhl$	15	$\frac{2!}{2 \cdot 0!} + 0 + 0 + \frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} +$ $0 + [(1.2.1) + (1.2.1) +$ $(2.1.1.1) + (2.1.1.1) +$ $(2.1.2.1)] = 15$
CVG-18 	$ab \quad ef \quad gh$ $(cde) \quad (cdf)$ $ijg \quad ijh$ $acki \quad bcki$ $adlj \quad bdlj$ $(ekgl) \quad fkgi$ $(ekhl) \quad fkhl$	11 + $\mu^{ced}$	$\frac{2!}{2 \cdot 0!} + 0 + 0 + \frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} +$ $0 + 0 + 0 + 0 + [(0.1.0) +$ $(1.2.1) + (2.1.1.1) +$ $(2.1.1.1) + (1.1.2.1)] = 11$
CVG-19 	$ab \quad ef \quad gh$ $cde \quad (cdf)$ $(ijg) \quad ijh$ $acki \quad bcki$ $adlj \quad bdlj$ $(ekgl) \quad (fkgi)$ $(ekhl) \quad (fkhl)$	9 + $\mu^{fkgi}$	$\frac{2!}{2 \cdot 0!} + 0 + 0 + \frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} +$ $0 + 0 + 0 + 0 + [(1.1.1) +$ $(1.1.1) + (2.1.1.1) +$ $(2.1.1.1) + (1.0.1.0)] = 9$
CVG-20 	$ab \quad ef \quad gh$ $(cde) \quad (cdf)$ $(ijg) \quad (ijh)$ $acki \quad bcki$ $adlj \quad bdlj$ $(ekgl) \quad fkgi$ $(ekhl) \quad (fkhl)$	8 + $\mu^{ced} +$ $\mu^{ijh}$	$\frac{2!}{2 \cdot 0!} + 0 + 0 + \frac{2!}{2 \cdot 0!} + \frac{2!}{2 \cdot 0!} +$ $0 + 0 + 0 + 0 + [(0.1.0) +$ $(0.1.0) + (2.1.1.1) +$ $(2.1.1.1) + (1.1.1.1)] = 8$

Chemie mbH), whose staff also assisted by means of many lively discussions. In addition, guidance on polytopal theory was gratefully received from Dr. R. J. Cook, Department of Pure Mathematics, University of Sheffield. Many thanks also to Dr. Cheng Qian, Chemical Abstracts Service, for Figure 16 and for last-minute mathematical insights!

## APPENDIX A

A comparative tabulation of ring sets for the DBR database (based on a tabulation by Nickelsen<sup>32</sup>) is given in Table III. Block 1 gives the DBR structure with its number and nullity. Block 2 gives the cycle sizes and types present. Block 3 gives the comparative perception of type for each ring set. Block 4 gives the number of rings found of each type. Block 5 gives comments about particular types, where necessary. Other symbols are  $\mu$  = nullity,  $G$  = ring size,  $e$  = simple cycle,  $\beta$  =  $\beta$ -ring,  $\mathcal{S}$  = SSSR,  $\mathcal{H}$  =  $\mathcal{H}$ -ring,  $\mathcal{E}$  = ESSR,  $\checkmark$  = included in ring set, and  $\times$  = not included in ring set. The comment codes given in the note column are  $F$  = simple face,  $A$  = SSSR has to choose arbitrarily between symmetrical equivalents,  $D$  = Doppelpunkte exclude cycle as a simple cycle,  $I$  = sym-

metrically equivalent maximum infinite region (simple face),  $M$  = maximal infinite region (simple face),  $N$  = Nachbarpunkte exclude cycle as a simple cycle,  $O$  = nonmaximal infinite region (simple face),  $P$  = primary cut face,  $R$  = alternative embedment region (simple face),  $S$  = secondary cut face, and  $Y$  = symmetrically equivalent alternative embedment finite region (simple face).

## APPENDIX B

Twenty representative cut-vertex graphs and their simple-face combinatorics are shown in Table IV.

Each simple face can be constructed by taking sequences of two or more of the alphabetically labeled arcs to form all possible regions of the cut-vertex graph. No distinction is made between non-CVP Nachbarpunkte single edges and the more general arcs. The lack of this distinction does not affect the combinatorial relationships. All pairs, triples, etc. of arcs under consideration are tabulated; those that cannot be topologically transformed into the infinite region of any planar projection are cut faces of the structure and are given in parentheses to show their exclusion as simple faces of the structure.

For CVG-1-10 all ULP and LPR arcs are labeled. This distinguishes the intra-CVP combinations from the inter-CVP combinations in the arc combination arrays. The intra-CVP combinations are always pairs of labeled arcs, while the inter-CVP combinations are triples, quadruples, etc. The inter-CVP combinations can become very long, and so for CVG-11-20 only those arcs that make some contribution to the combinations are labeled. It is consequently harder to distinguish between the intra- and inter-CVP combinations in these arrays. In addition, many of the arrays have been compacted by moving elements around to enable the use of fewer columns.

For the combinatorial equations, the intra-CVP combinations are given first followed by the inter-CVP combinations enclosed in square brackets. Those arcs that contribute nothing to the combinations are indicated by a zero.

The column marked  $\mathcal{F}$  is the total number of simple faces for each cut-vertex graph and includes the nullity ( $\mu$ ) number of rings for each LPR where appropriate. In contrast, the combinatorial equation only includes the non-LPR simple faces.

### BIBLIOGRAPHY

- (1) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Review of Ring Perception Algorithms for Chemical Graphs. *J. Chem. Inf. Comput. Sci.* (first of four papers in this issue).
- (2) Downs, G. M. Computer storage and retrieval of generic structures in patents: ring perception and screening to extend the search capabilities. Ph.D. Thesis, University of Sheffield, March 1988; Chapter 4 (Theoretical considerations of ring perception).
- (3) Elk, S. B. Effect of Taxonomy Class and Spanning Set on Identifying and Counting Rings in a Compound. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 11-16.
- (4) Zamora, A. An Algorithm for Finding the Smallest Set of Smallest Rings. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 40-43.
- (5) Lederberg, J. DENDRAL-64: a system for computer construction, enumeration and notation of organic molecules as tree structures and cyclic graphs. Part 1: notational algorithm for tree structures. Report CR-57029; National Aeronautics and Space Administration: Washington, DC, 1969.
- (6) Carhart, R. E.; Smith, D. H.; Brown, H.; Sridharan, N. S. Applications of Artificial Intelligence for Chemical Inference. XVI. Computer Generation of Vertex-Graphs and Ring Systems. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 124-130.
- (7) Sridharan, N. S. Computer generation of vertex graphs. *Inf. Process. Lett.* **1974**, *3*(2), 57-63.
- (8) Balaban, A. T.; Filip, P.; Balaban, T. S. Computer program for finding all possible cycles in graphs. *J. Comput. Chem.* **1985**, *6*, 316-329.
- (9) Adamson, G. W.; Cowell, J.; Lynch, M. F.; Town, W. G.; Yapp, A. M. Analysis of structural characteristics of chemical compounds in a large computer-based file. Part 4. Cyclic fragments. *J. Chem. Soc., Perkin Trans. 1* **1973**, 863-865.
- (10) Adamson, G. W.; Creasey, S. E.; Eakins, J. P.; Lynch, M. F. Analysis of structural characteristics of chemical compounds in a large computer-based file. Part 5. More detailed cyclic fragments. *J. Chem. Soc., Perkin Trans. 1* **1973**, 2071-2076.
- (11) Mawby, G. A preliminary evaluation of the generic structure search system, GENSEARCH, designed for use in chemical patent searching. M.Sc. Dissertation, University of Sheffield, Sept 1984.
- (12) Kirk, S. Generic Chemical Structures Project. Evaluation of the search system, Part 3: the effect of ring screens on the performance of the fragment screening applied to structures in the specific and level-3 databases. M.Sc. Dissertation, University of Sheffield, Sept 1986.
- (13) Behzad, M.; Chartrand, G. *Introduction to the theory of graphs*; Allyn & Bacon: Boston, 1971.
- (14) Behzad, M.; Chartrand, G.; Lesnick-Foster, L. *Graphs and digraphs*; Prindle, Weber & Schmidt: Boston, 1979.
- (15) Berge, C. *Graphs and hypergraphs*; North-Holland: Amsterdam, 1973; Mathematical Library 6.
- (16) Mayeda, W. *Graph theory*; Wiley-Interscience: New York, 1972.
- (17) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1972.
- (18) Marshall, C. *Applied graph theory*; Wiley-Interscience: New York, 1971.
- (19) Bondy, J. A.; Murty, U. S. R. *Graph theory with applications*; Macmillan: London, 1976.
- (20) Deo, N. *Graph theory with applications to engineering and computer science*; Series in Automatic Computation; Prentice-Hall: Englewood Cliffs, NJ, 1974.
- (21) Balaban, A. T., Ed. *Chemical applications of graph theory*; Academic Press: London, 1966.
- (22) Trinajstić, N. *Chemical graph theory*; CRC Press: Boca Raton, FL, 1983; 2 vol.
- (23) Jordan, C. Sur les assemblages de lignes. *Angew. Math.* **1869**, *70*, 185-190; *Oeuvres* **4**, 303-308.
- (24) Cauchy, A. L. Recherches sur les polyèdres—premier mémoire. *J. École Polytech.* **1813**, *9*(16), 68-86; *Oeuvres* **2**, *1*, 7-25.
- (25) Euler, L. Elementa doctrinae solidorum. *Novi. Comm. Acad. Sci. Imp. Petropol.* **1752-3**, *4*, 109-140; *Opera Omnia* **1**, *26*, 72-93.
- (26) Schlegel, V. Theorie der homogen zusammengesetzten Raumgebilde. *Nova Acta Leopold.* **1983**, *44*, 344-459.
- (27) 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*, 204-206.
- (28) Courant, R.; Robbins, H. *What is mathematics?—an elementary approach to ideas and methods*; Oxford University Press: Oxford, U.K., 1941.
- (29) Grünbaum, B. *Convex polytopes*; Wiley-Interscience: New York, 1967.
- (30) Balinski, M. On the graph structure of convex polyhedra in n-space. *Pac. J. Math.* **1961**, *11*, 431-434.
- (31) Steinitz, E. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.*, **44**, 1-40.
- (32) Fugmann, R.; Dölling, U.; Nickelsen, H. A topological approach to the problem of ring structures. *Angew. Chem., Int. Ed. Engl.* **1967**, *6*, 723-733.
- (33) Plotkin, M. Mathematical Basis of Ring-Finding Algorithms in CIDS. *J. Chem. Doc.* **1971**, *11*, 60-63.
- (34) Nickelsen, H. Ringbegriffe in der Chemie-Dokumentation. *Nachr. Dok.* **1971**, *3*, 121-123 (and associated microfiche).
- (35) Fujita, S. Logical Perception of Ring Opening, Ring Closure, and Rearrangement Reactions Based on Imaginary Transition Structures. Selection of the Essential Set of Essential Rings (ESER). *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 1-9.
- (36) Fujita, S. A New Algorithm for Selection of Synthetically Important Rings. The Essential Set of Essential Rings for Organic Structures. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 78-82.
- (37) Wipke, W. T.; Dyott, T. Use of Ring Assemblies in a Ring Perception Algorithm. *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 140-144.
- (38) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Computer Storage and Retrieval of Generic Chemical Structures in Patents. 9. An Algorithm To Find the Extended Set of Smallest Rings in Structurally Explicit Generics. *J. Chem. Inf. Comput. Sci.* (third of four papers in this issue).
- (39) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Computer Storage and Retrieval of Generic Chemical Structures in Patents. 10. Assignment and Logical Bubble-Up of Ring Screens for Structurally Explicit Generics. *J. Chem. Inf. Comput. Sci.* (fourth of four papers in this issue).