An Algorithm for Machine Perception of Synthetically Significant Rings in Complex Cyclic Organic Structures

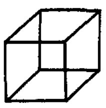
E. J. Corey* and George A. Petersson

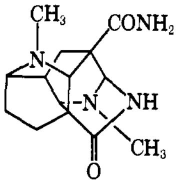
Contribution from the Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *Received January 30, 1971*

Abstract: Computer-assisted synthetic analysis requires a systematic method of recognizing rings in polycyclic structures. This method should be generally applicable, fast, and should provide a concise description of the cyclic structure. The approach that has been developed operates on an n-cyclic structure (containing as many as $2^n - 1$ rings) to select a "synthetic subset" of rings which provide a concise but complete description. The description is complete in the sense that any ring not in the "synthetic subset" can be generated by taking the logical *exclusive or* of the sets of bonds in rings which have been included in the "synthetic subset." This subset is defined as the set of all rings which cannot be expressed as the *exclusive or* of smaller rings, plus any other rings not larger than six. Algorithms for efficiently obtaining this "synthetic subset" have been developed, and results from applying these algorithms to a variety of ring networks are presented.

Chemists concerned with the design of synthetic processes generally analyze cyclic organic structures by examining either a three-dimensional model or a two-dimensional drawing and considering whether certain of the rings can be constructed by known methods. There is a tendency to look for familiar structural patterns rather than to trace the network systematically and exhaustively. This purely intuitive approach is adequate for simple cyclic systems but may fail to recognize important features of more complex ring networks unless a systematic procedure is introduced. Computer programs for synthetic analysis[1,2] and organic structure or substructure search[3,4] require a systematic approach for even the simplest cyclic systems, since computers lack intuition. This approach should provide a complete description of a general ring network. It should be fast, and the description should be concise, to reduce the time required for its use or interpretation.

The most obvious procedure to guarantee the perception of all *important* rings is clearly one in which all rings are perceived. This is not very efficient for large systems, since the total number of rings in an n-cyclic system (the cyclic order of a structure is given by the number of bonds minus the number of atoms plus one) can be as large as $2^n - 1$, and usually only about n of these will be of interest to a synthetic chemist. For example, the pentacyclic molecules cubane (I) and N-methyldihydronicotinamide dimer (II) contain a total of 28 and 29 rings, respectively. In the case of cubane, only the six four-membered rings are of chemical interest (see below), and in the case of dinicotinamide, only the nine six-membered rings are of synthetic significance. Therefore, it is desirable to consider only the subset of rings which is meaningfully related to the synthetic chemistry of the system. This subset will be designated in the discussion which follows as a "synthetic subset." Two attempts to solve this problem have already been published.[1,5] We shall compare these with the present approach in a later section of this paper.


I



II

Generally the formation of a bond within a ring network depends most critically on the size of the smallest ring in which that bond is contained (neglecting considerations of functional groups). Therefore, it is required that the synthetic subset include a ring if, for some bond, there is no smaller ring containing that bond. For the purposes of synthetic analysis, larger rings are sometimes important (*e.g.*, the six-membered ring in norbornane is critical for synthesis *via* Diels–Alder addition). If the synthetic subset of rings provides a complete description of the ring system, it will be possible to treat these larger rings implicitly rather than including them in the subset which is listed explicitly. In order to define what is meant by a complete description of the ring system, it is necessary to consider some basic concepts of graph theory.

Background. As is well known, group theory is a powerful tool for dealing with molecular symmetry. One of the most elementary applications of group theory makes use of the closure property of groups to test a list of symmetry operations for completeness (*i.e.*, to check whether all products of symmetry operations are already included in the list of symmetry operations). If a molecule has many symmetry elements, we may start with certain partial lists and generate the remainder by taking products of those we already have. Similarly, a ring operation with a closure property can be useful for ring perception.

Consider the bonds in either one of the two six-membered rings of decalin, but not the bond which appears in both. These bonds form the ten-membered envelope ring. We can describe the bicyclic molecule, decalin, as consisting of two six-membered rings. This is a complete description, since the ten-membered

---

(1) E. J. Corey and W. T. Wipke, *Science*, **166**, 178 (1969).

(2) E. J. Corey, *Quart. Rev., Chem. Soc.*, **25**, No. 4 (1971).

(3) H. L. Morgan, *J. Chem. Doc.*, **5**, 107 (1965).

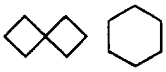
(4) E. H. Sussenguth, Jr., *ibid.*, **5**, 36 (1965). See also J. Lederberg, *Proc. Nat. Acad. Sci. U. S.*, **53**, 134 (1965).

(5) R. Fugmann, U. Dölling, and H. Nickelsen, *Angew. Chem., Int. Ed. Engl.*, **6**, 723 (1967).

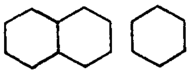
peripheral ring can be generated by taking the bonds which appear in either of the two six-membered rings, but not in both. Its existence is implied by the statement that decalin consists of two six-membered rings with one bond in common. The logical *exclusive or* of the bonds is the ring operation we seek.

We shall now proceed to a more formal approach which will enable us to make full use of the groups and vector spaces associated with this operation. We shall restate some fundamental theorems of graph theory[6] in the conventional terms used to discuss chemical structures. It is useful first to define some terms which are not in general use by chemists. The first time a chemical term appears, the corresponding term from graph theory[6] will be given in parentheses.

The *connectivity* of an atom (vertex) is the number of atoms directly bonded to it (adjacent to it). A *tree* is a connected acyclic structure (graph). A *spanning tree* contains all the atoms in a structure. A *ring closure bond*[3] (chord or link) relative to a given spanning tree is a bond not in the spanning tree. A *ring* (elementary circuit containing three or more vertices) is a closed path through three or more atoms, none of which appears in the path more than once. When we refer to a ring, we shall generally mean the set of bonds in the ring. A *bond disjoint union of rings* is the set of all bonds appearing in two or more rings, no pair of which is bridged or fused (*e.g.*, the set of bonds indicated in III is a bond disjoint union of rings, but the set of bonds indicated in IV in not). If $\mathbf{S}_1$ and $\mathbf{S}_2$ are two sets of bonds, then their ring sum, $\mathbf{S}_1 \oplus \mathbf{S}_2$, is the set of bonds in either $\mathbf{S}_1$ or $\mathbf{S}_2$, but not in both (*i.e.*, the ring sum is equivalent to the logical *exclusive or*).



III



IV

Suppose the ring sum, $\mathbf{S}$, is taken of two rings, $\mathbf{R}_1$ and $\mathbf{R}_2$. If $\mathbf{R}_1$ equals $\mathbf{R}_2$, then $\mathbf{S}$ is the empty set, ϕ. Otherwise, $\mathbf{S}$ is not empty. If $\mathbf{S}$ is not empty, we can start at an arbitrary atom, a_i, in $\mathbf{S}$ and begin tracing a path through the bonds $(a_i, a_j)(a_j, a_k)$... such that no bond appears in the path more than once. Since the connectivity of any atom in $\mathbf{R}_1$ or $\mathbf{R}_2$ is two, the connectivity of any atom in $\mathbf{S}$ is either two ($\mathbf{R}_1$ and $\mathbf{R}_2$ are bridged, fused, or isolated) or four ($\mathbf{R}_1$ is spiro joined to $\mathbf{R}_2$). Therefore, whenever we arrive at an atom, a_l, through bond (a_m, a_l), we can leave through the bond (a_l, a_n) until we return to atom a_i. If there are atoms in $\mathbf{S}$ which are not on this path, we can repeat this procedure. The set $\mathbf{S}$ is therefore either a ring or a bond disjoint union of rings if we consider ϕ to be a "ring." Furthermore, the set of all rings and bond disjoint unions of rings is closed under the operation ring sum.

The set of all rings and bond disjoint unions of rings in any molecule forms a group under the operation ring sum (ϕ is the identity element and each ring is its own inverse). Together with the scalar field of integers modulo 2 (*i.e.*, the numbers zero and one), they form a vector space. Each ring or bond disjoint union of rings is a vector in this space. The vector space is closed under addition (ring sum) and scalar multiplication ($0 \cdot \mathbf{R} = \phi$, and $1 \cdot \mathbf{R} = \mathbf{R}$). We shall now examine some properties of this ring space.

Let n_a be the number of atoms and n_b the number of bonds in a structure. There are $n_a - 1$ bonds in a spanning tree. Relative to any spanning tree there are then $n_b - n_a + 1$ ring closure bonds. Suppose one of these ring closure bonds joins atoms a_i and a_j. The spanning tree contains these atoms (by definition) and a path between them (a tree is connected). This path is unique, since a tree is acyclic. If we add the ring closure bond to the spanning tree, we therefore form one ring. This is called a *fundamental ring* (fundamental circuit).[7] There is a one-to-one correspondence between the set of $n_b - n_a + 1$ ring closure bonds and this *fundamental system of rings* relative to any spanning tree.

Let $\mathbf{R} = \{b_1, b_2, ..., b_i, b_{i+1}, b_{i+2}, ..., b_j\}$ be a ring, where $b_1, b_2, ...,$ and b_i are ring closure bonds and $b_{i+1}, b_{i+2}, ...,$ and b_j are bonds in the associated spanning tree. Let $\mathbf{R}_k$ be the fundamental ring corresponding to ring closure bond b_k, and let $\mathbf{R}^1 = \mathbf{R}_1 \oplus \mathbf{R}_2 \oplus ... \oplus \mathbf{R}_i$. Since $\mathbf{R}$ and $\mathbf{R}^1$ both contain bonds $b_1, b_2, ..., b_i$, $\mathbf{R} \oplus \mathbf{R}^1$ contains only bonds in the spanning tree. But $\mathbf{R} \oplus \mathbf{R}^1$ is a ring or a bond disjoint union of rings. The two previous statements are contradictory unless $\mathbf{R} \oplus \mathbf{R}^1 = \phi$ so that $\mathbf{R} = \mathbf{R}^1$. Therefore, any ring can be expressed as a linear combination of the fundamental rings for a given spanning tree. Since $\mathbf{R}_k$ is the only fundamental ring containing the bond b_k, the fundamental rings are linearly independent (*i.e.*, $\mathbf{R}_k$ cannot be expressed as a ring sum of the other fundamental rings) and therefore form a basis for the ring space. The dimension of the ring space is equal to the cyclic order, $n_b - n_a + 1$. Including the empty set, ϕ, there are $2^{n_b - n_a + 1}$ possible linear combinations of these basis vectors and therefore a maximum of $2^{n_b - n_a + 1} - 1$ rings. Several algorithms have been published,[8] which generate all rings from a set of fundamental rings by taking the ring sum. Even when efficiently programmed, the required generation of all $2^n - 1$ rings and bond disjoint unions of rings is very time consuming. In addition, one is faced with the problem of storing the resulting long ring lists and scanning them in future use.

Theory. We can now make our intuitive concept of a complete description of a ring system more concrete. We wish the synthetic subset of rings to span the ring space (*i.e.*, they must *include* a basis for the ring space). This guarantees that our subset will provide a complete description of the ring network in the sense that any ring not in our subset can be generated by taking the ring sum of rings in our subset. It is not necessary that the rings in our subset be linearly independent. It is clearly desirable to include all six four-membered rings for cubane (I). Since any five of these rings form a basis for the cubane ring space, the set of all six is linearly dependent.

We can now define the set of *minimum spanning rings* for any structure. A minimum spanning ring is a

---

(6) C. L. Liu, "Introduction to Combinatorial Analysis," McGraw-Hill, New York, N. Y., (1968).

(7) We are adopting the notation of reference 6, not that of reference 5.

(8) (a) P. L. Long, R. F. Phares, J. E. Rush, and L. J. White, abstract CHLT15, 160th National Meeting of the American Chemical Society, Chicago, Ill., Sept 1970; (b) N. E. Gibbs, *J. Ass. Computing Mach.*, **16**, 564 (1969).

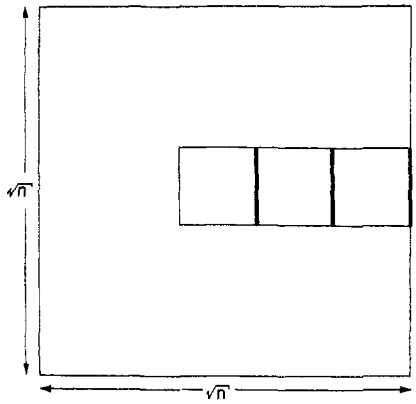


Figure 1. A square network of cyclic order n has sides which include the bonds from $n^{1/2}$ rings. Three ring closure bonds are indicated by broader lines. The associated four-, six,- and eight-membered fundamental rings "penetrate" to a depth of one, two, and three rings, respectively.

ring which cannot be expressed as a ring sum of smaller rings.

The set of minimum spanning rings includes the rings of primary synthetic interest. If for some bond, b, there is no ring smaller than **R** containing that bond, then **R** is a minimum spanning ring, since it cannot be expressed as a ring sum of smaller rings. This definition also guarantees that our synthetic set will include a basis for the ring space. We shall demonstrate this by assuming that it is not true and arriving at a contradiction. Suppose some ring, **R**, were linearly independent of the set of minimum spanning rings. Then **R** is not a minimum spanning ring, and it follows from the definition that **R** can be expressed as a linear combination of rings smaller than **R** ($\mathbf{R} = \mathbf{R}_1 \oplus \mathbf{R}_2 \oplus \ldots \oplus \mathbf{R}_i$). Since **R** is linearly independent of the set of minimum spanning rings, so is one of these smaller rings, $\mathbf{R}_j$ ($1 \leq j \leq i$). If we iterate this process, we shall eventually conclude that if **R** is linearly independent of the set of minimum spanning rings, then so is some ring, $\mathbf{R}_k$, which is smaller than all rings. Since this conclusion is impossible, the set of minimum spanning rings must include a basis for the ring space. This can be important. The pentacyclic structure, V,

V

contains four three-membered rings, each of which is the smallest ring for some bond. In order to include the obviously present four-membered ring, we must require that we include a basis for the ring space in our set. Otherwise, we shall not distinguish between structures V and VI.

VI

The definition of minimum spanning rings requires that any such ring will either be synthetically significant or will be necessary for a complete description of the ring system. The inclusion of a basis implies that the set of minimum spanning rings contains a minimum of $n_b - n_a + 1$ rings. The number of minimum spanning rings will frequently equal the cyclic order of a system. Hence, the set of minimum spanning rings provides a complete and concise description of a ring network.

We shall now present a method for finding the minimum spanning rings for an arbitrary ring network. Suppose we have a set of rings, $\mathbf{S}_r$, which includes the set of minimum spanning rings. Let $\mathbf{S}_{ms}$ be an empty set. If we keep adding the smallest ring(s) in $\mathbf{S}_r$ which is linearly independent of the rings in $\mathbf{S}_{ms}$ to $\mathbf{S}_{ms}$ until no such ring exists, then $\mathbf{S}_{ms}$ will be the set of minimum spanning rings.

The efficiency of our approach depends upon the generation of an appropriate set of rings, $\mathbf{S}_r$, which contains little besides the set of minimum spanning rings. The following property of minimum spanning rings will be useful in this context. Let $\mathbf{R} = \{b_1, b_2, \ldots, b_i, b_{i+1}, b_{i+2}, \ldots, b_j\}$ be a ring, where $b_1, b_2, \ldots, b_i$ are ring closure bonds. Let $\mathbf{R}_k$ be the fundamental ring for bond b_k. Suppose **R** is larger than all the $\mathbf{R}_k$ ($k = 1, 2, \ldots, i$). Then **R** cannot be a minimum spanning ring, since it can be expressed as a linear combination of rings smaller than **R**. Therefore, if for every ring closure bond, b_l, we collect all rings not larger than $\mathbf{R}_l$, which contain b_l, we shall include all minimum spanning rings in our collection.

Consider the set of fundamental rings, $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_n\}$, relative to some spanning tree. If the ring, $\mathbf{R}_\alpha = \mathbf{R}_i \oplus R_j$, is smaller than either $\mathbf{R}_i$ or $\mathbf{R}_j$, we can replace this fundamental ring with $\mathbf{R}_\alpha$ and still have a basis. If we iterate this process to convergence, we shall have a *reduced basis*, $\{\mathbf{R}_{\alpha_1}, \mathbf{R}_{\alpha_2}, \ldots, \mathbf{R}_{\alpha_n}\}$. Let $\mathbf{R}_{\alpha_{max}}$ be the largest ring in a reduced basis. If follows from the definition of minimum spanning rings that if for every ring closure bond, b_l, we collect all rings not larger than *either* $\mathbf{R}_l$ or $\mathbf{R}_{\alpha_{max}}$, we shall include all minimum spanning rings in our collection. If the spanning tree is grown in an appropriate fashion,[9] the reduced basis will frequently be the set of minimum spanning rings.

In order to compare this approach with previous methods,[1,5] it is necessary to estimate the number of rings that we must collect. Unlike algorithms for numerical computations, the number of steps we must perform for ring perception depends upon the details of the data (*i.e.*, the ring network). Therefore, our timing estimates will necessarily be rather crude.

The number of rings we must collect is proportional to the number of ring closure bonds, n, and the number of rings found for each ring closure bond. In a linear ring network (*e.g.*, the homologous series; benzene, naphthalene, anthracene, etc.) only one ring must be found for each ring closure bond. In a square network the number of rings found for a ring closure bond will vary according to how deeply the fundamental ring "penetrates" the network (Figure 1). The worst case would essentially require analysis of a ring network of cyclic order $n^{1/2}$ giving a total of $2^{n^{1/2}} - 1$ rings. The number of rings found for an "average" ring closure bond in a planar[10] network should then be approximately $1/2 \times 2^{n^{1/2}}$. Similarly, the number of rings found for each ring closure bond in a nonplanar network is approximately $2^{(n^{1/2}-1)}$. The total number

---

[9] K. Paton, *Commun. Ass. Computing Mach.*, 12, 514 (1969).

[10] A planar graph is one that *can* be drawn so that the bonds intersect *only* at the atoms they join.

of rings we must collect to analyze an *n*-cyclic structure is approximately given by

$$N(n) \equiv \left\{\begin{matrix} n^{1/2}\text{ linear structure} \\ n2^{(n^{1/2}-1)}\text{ planar structure} \\ n2^{(n^{1/3}-1)};\text{ nonplanar structure} \end{matrix}\right\} \qquad (1)$$

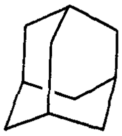
when the above approach to obtaining the set of minimum spanning rings is used. This is really an upper bound, since we have not taken into account the use of a reduced basis. The number of rings we must collect will be considerably smaller than the total number of rings.

An alternative approach to obtaining a set of rings which includes all minimum spanning rings would be to take linear combinations of the fundamental rings relative to some spanning tree. This is analogous to our earlier discussion of the symmetry operations for a molecule. If m is the maximum number of ring closure bonds in any minimum spanning ring, we need only take linear combinations of m or fewer fundamental rings ($\mathbf{R}_k = \mathbf{R}_{k1} + \mathbf{R}_{k2} + \ldots + \mathbf{R}_{km}$). There are a total of $\sum_{j=1}^{m} n!/(n - j)!\, j!$ such linear combinations. If M is the size of the largest minimum spanning ring, then clearly m is less than or equal to the minimum of M and n. We could determine M by initially assuming it is 3 and then increasing it by 1 until we find a set of minimum spanning rings, none of which contains more than M bonds. The number of linear combinations we must then consider is approximately $n^M/M!$. Although the operation ring sum is very rapid, the number of times we must perform this operation increases much more rapidly than the number of rings we must grow with the above algorithm, for complex molecules of possible interest. The number of intermediate quantities we must store also grows much more rapidly. If we used a value of M smaller than the number of bonds in the largest minimum spanning ring, we have no guarantee that we will include the set of minimum spanning rings in our collection.

Two noteworthy attempts have previously been made to define chemically important rings. The first of these[5] gives results similar to the present approach, but the definition and the resulting algorithm are extremely complex (all possible paths between all possible pairs of bridgehead atoms must be found). The complex definition makes it difficult to get an intuitive grasp of which rings will be perceived (or to derive the general properties of the ring lists produced).

The number of bridgehead atoms in an *n*-cyclic structure can be as large as $2(n - 1)$. We shall assume an "average" number of bridgehead atoms equal to n. There are then $n(n - 1)/2$ pairs of bridgehead atoms. Let atoms a_1 and a_2 be one such pair. Let a_j be any other bridgehead atom with connectivity c_j. If on a path from a_1 to a_2 we arrive at atom a_j, there are then $c_j - 1$ paths through which we could leave a_j, other than the path we arrived on. If c is the "average" connectivity, there are approximately $(c - 1)^n$ elementary paths[11] between each pair of bridgehead atoms. The total number of elementary paths between all pairs of bridgehead atoms is approximately $n(n - 1)(c - 1)^n/2$. The number of paths grows far more rapidly than the total number of rings. There are 384 elementary paths between pairs of bridgehead atoms in cubane (the number predicted by the above formula is $5 \cdot 4 \cdot 2^5/2 = 320$). Clearly this approach is not very practical for complex ring networks.

The very large increase in computing time over the present approach does not seem to be justified. The basic difference in the resulting ring sets is that the previous algorithm will recognize all three rings defined by a bridged network [*e.g.*, it will recognize four six-membered rings and three eight-membered rings in adamantane (VII)]. The set of minimum spanning rings includes only the two smallest rings defined by a bridged network [*e.g.*, it includes only the four six-membered rings in adamantane (VII)]. Special bicyclic structures such as norbornane and indole which are strategically important for synthetic analysis can easily be perceived from the minimum spanning rings (*e.g.*, if the ring sum of two five-membered rings is a six-membered ring, a norbornane substructure is present). The earlier definition of important rings[5] is very time consuming to apply and includes rings which are of no synthetic interest.



VII

The second definition of chemically important rings is more tractable, but it too has serious disadvantages. The algorithm for obtaining "real" rings[1] begins by the direct perception of all rings. It is therefore considerably slower than the present approach, although not as slow as the other earlier algorithm,[5] when applied to complex ring networks. The set of "real" rings may include rings of questionable interest (*e.g.*, four six-membered rings are "real" rings in cubane), and there is a discrepancy between the definition of "real" rings and the algorithm given for obtaining them (no *maximum proper covering set*[1] of five rings exists for the pentacyclic structure, V).

Both of the earlier approaches stress a correspondence to the rings intuitively recognized by the chemist. This is a rather vague criterion. We believe that the requirement that we include a basis for the ring space and any ring which is the smallest ring for some bond provides a more solid foundation for ring perception.

It is possible that for some purposes our set of chemically important rings should include all rings containing six or fewer atoms. This is readily accomplished by continuing to seek rings containing a given ring closure bond until we have both included the fundamental ring for this bond and grown all rings containing six or fewer atoms. The resulting set of rings may be considerably larger and the computing time required may increase substantially. Ultimately the set of chemically important rings one collects must depend upon the use one wishes to make of them. Although one may wish to include additional rings for some uses, any description of a molecular ring network should include the set of minimum spanning rings.

A Ring Perception Algorithm. The initial information used in the perception of rings consists of the atom and bond tables, routines for extracting data from the

---

(11) An elementary path is one which does not contain any atom more than once.

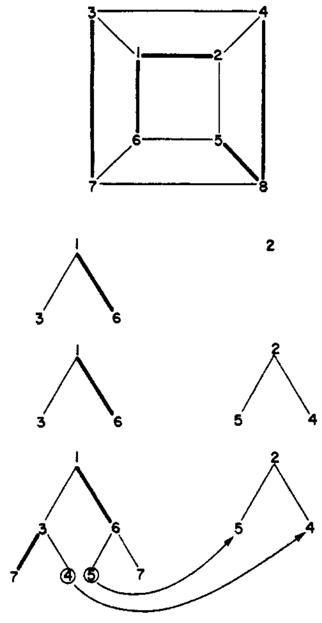


Figure 2. Rings are formed by growing a tree from each end of a ring closure bond.

atom and bond tables, and certain primary sets which result from the operation of the perception executive on the atom and bond tables. Atoms which are not in rings could be deleted in a first step. However, this is handled automatically by the tree-growing process of Paton[9] which yields a spanning tree, $\mathbf{S}_t$. For each atom in the tree, we store the set of bonds above it in the tree. When an atom is encountered that is already in the spanning tree, a fundamental ring is obtained by taking the ring sum of the two bond sets. This requires $n_b \cdot n_a$ bits of additional storage, but eliminates the necessity of "walking up the tree." We then form the corresponding reduced basis by retaining the smallest two rings of each triplet; $\mathbf{R}_i$, $\mathbf{R}_j$, $\mathbf{R}_i \oplus \mathbf{R}_j$; where $i > j$ and j varies from 1 to $n - 1$ where n is the number of fundamental rings.

Now for each ring closure bond, b_l, we must find all rings containing b_l which are not larger than either the fundamental ring containing b_l or the largest reduced basis ring. We do this by growing a tree from each end of b_l. For example, suppose we have the cubane spanning tree indicated in Figure 2, and the largest reduced basis ring is a four-membered ring. The trees for ring closure bond (1,2) are indicated above. They are grown one level at a time, alternating between the trees for atoms one and two. It is only necessary to compare a newly grown level with the most recent level on the other side to perceive rings. Rather than store the entire tree, we need only store the set of bonds above each atom in the current level. When a match is found between the two sides, the ring sum of the two bond sets gives the ring in question. This process is iterated until either the fundamental ring is found or the size of the largest reduced basis ring is reached.

These rings are stored as bond sets and are used to select the set of minimum spanning rings. The union of the bond sets of the fundamental system of rings is the set of all ring bonds.

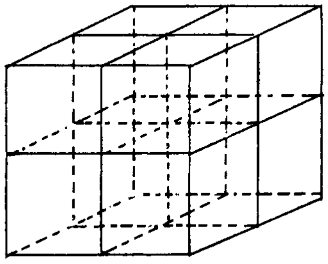
We now select the smallest rings and include them in the set of minimum spanning rings, $\{\mathbf{R}_{ms}\}$. If the set of minimum spanning rings does not cover all ring bonds, we consider the set of smallest rings not yet examined. Any of these rings which include bonds not previously covered are then included in the set of minimum spanning rings. This process is iterated until all ring bonds are covered.

Next, we construct the set of ring closure bonds whose fundamental ring has not been included in the set of minimum spanning rings, $\mathbf{B}_0$.

If some ring in $\{\mathbf{R}_{ms}\}$ contains only one bond in $\mathbf{B}_0$, we remove that bond from $\mathbf{B}_0$. This process is iterated until a pass does not remove any bonds from $\mathbf{B}_0$. If there are still bonds remaining in $\mathbf{B}_0$, we find the rings in $\{\mathbf{R}_{ms}\}$ which contain bonds in $\mathbf{B}_0$. (These rings exist, since $\{\mathbf{R}_{ms}\}$ is a covering set.) We now form the set of rings $\{\mathbf{R}_{lc}\}$ which are linear combinations of these rings. If any ring in $\{\mathbf{R}_{lc}\}$ contains only one bond in $\mathbf{B}_0$, we remove that bond from $\mathbf{B}_0$. If there are still bonds remaining in $\mathbf{B}_0$, we find the smallest ring(s) $\{\mathbf{R}_s\}$ which contains bond(s) in $\mathbf{B}_0$ but is (are) not in $\{\mathbf{R}_{lc}\}$, and add them to $\{\mathbf{R}_{ms}\}$. The entire preceding paragraph is iterated until no bonds remain in $\mathbf{B}_0$.

Results

A FORTRAN IV version of the above algorithm has been applied to a variety of ring networks, and the results are summarized in Table I. The total number of rings in a network grows very rapidly with the cyclic order, n, and can be quite close to the upper bound of $2^n - 1$. The number of minimum spanning rings, however, is never much larger than the cyclic order, n (Table I). The approximate upper bound for the number of rings we must collect if we do not use a reduced basis, $N(n)$, is surprisingly accurate for all examples in Table I. In particular, we note that $N(n)$ correctly predicts that the number of rings we must grow will be considerably smaller for the nonplanar[10] structure VIII than for the planar[10] icosahedron, even though the cyclic order of VIII is much larger. When a reduced basis is used, *only* the minimum spanning rings are grown, except in the most complicated ring networks. Since the number of rings grown is the essential factor which determines the computing time, this time is also approximately linear with n for the simpler networks.



VIII

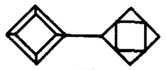
The algorithm discussed in this paper has two very desirable properties. The first is that if several simple ring networks are joined together so that no new rings are formed, the computing time for the larger system is approximately the sum of the computing times for the separate simple ring networks. The second desirable property is that the exponential dependence

Table I

Molecule	n	No. of ms rings	$2^n - 1$	No. of rings	$N(n)$[a]	No. of rings grown[b]	No. of rings grown[c]
Cyclohexane	1	1	1	1	1	1	1
Cholesterol	4	4	15	10	4	4	4
Dinicotinamide (II)	5	9	31	29	12	9	9
Cubane (I)	5	6	31	28	12	14	6
Cyclospirane (V)	5	5	31	20	12	12	5
Cyclospirocubane (IX)	10	11	1,023 (62)[e]	48	24[d]	26	11
Dodecahedrane	11	12	2,047	1,168	55	51	12
Icosahedron	19	20	524,287	12,878	196	208	208
Octacubane (VIII)	28	36	268,435,455	?	115	118	118

[a] Eq 1. [b] Without using a reduced basis. [c] Using a reduced basis. [d] Since this molecule represents the acyclic joining of two smaller networks, we take the sum of these functions for the smaller networks.

upon the cyclic order has been removed, permitting the analysis of complex ring networks.



IX

The set of minimum spanning rings is a small and therefore convenient subset of the rings, which completely defines the ring network. For some applications one may wish to include additional rings, but the set of minimum spanning rings should be a part of any description of a cyclic molecule.

Applications to Synthetic Analysis. The smallest ring in which a given bond appears places a special set of restrictions on how that bond can be formed, and this information is critical to the evaluation of the applicability of a synthetic reaction to the generation of a particular cyclic structure. The aldol reaction, for instance, is well suited to the construction of six- but not four-membered rings. The perception of rings in a structure plays a key role in the *selection* of suitable synthetic reactions as well as in their *evaluation.* For example, recognition of a six-membered ring is a first step in selecting synthetic processes such as the Diels–Alder, Birch reduction, cation–olefin cyclization, or Robinson annulation reaction. Perception of individual rings is also a prerequisite to the determination of critical relationships between rings in a complex network and the recognition of strategic bond disconnections which lead to maximum simplification of the network.

It appears that for the purpose of synthetic analysis it is desirable to recognize the set of minimum spanning rings plus any other rings with six or fewer bonds. This can be accomplished by the following approach. First we find the set of minimum spanning rings as described above. If the ring sum of any pair of these rings, or the ring sum of all of them, is not larger than a six-membered ring, we add it to the set of rings we consider for synthetic analysis.

In closing, we present a brief summary of the procedure for obtaining the set of minimum spanning rings. If one or more rings are present, then the following operations are carried out.

1. GROW A SPANNING TREE. FIND FUNDAMENTAL RINGS (FR): Encountering an atom already in the spanning tree indicates FR.

2. REMOVE ENVELOPE RINGS TO FORM REDUCED BASIS. For each triplet of rings, R_i, R_j, $R_i \oplus R_j$, retain two smallest rings.

3. FOR EACH RING CLOSURE BOND, b_c, FIND RINGS CONTAINING b_c NOT LARGER THAN FR OR LARGEST REDUCED BASIS RING. Grow a tree from each end of b_c; a common atom in the two trees indicates a ring; iterate until the smaller of (a) fundamental or (b) the largest reduced basis ring is found.

4. ORDER THESE RINGS BY SIZE AND STORE AS BOND SETS ($\cup$ FR $= \{$RING BONDS$\}$).

5. SELECT SMALLEST RING not in $\{$MSR$\}$ with bonds not in $\cup$ MSR and place them in $\{$MSR$\}$; iterate until $\cup$ MSR $= \cup$ FR.

Acknowledgments. The authors are grateful to the National Institutes of Health for financial assistance and to the Advanced Research Projects Agency for their support of the Center for Research in Computer Sciences at Harvard.