

## The Atom Connectivity Matrix (ACM) and Its Characteristic Polynomial (ACMCP)\*

By LEONARD SPIALTER

Chemistry Research Laboratory, ARL, Wright-Patterson Air Force Base, Ohio

Received February 24, 1964

### INTRODUCTION

Chemical pictographs or molecular diagrams are the natural language in which chemists describe molecules and reactions thereof. Unfortunately, such pictographs do not lend themselves to direct filing and comparison primarily because dimensions, orientation, and angles are not important fundamentals in the usual case. The essential components are the constituent atoms and the bonds between them (in the current view of molecular structure). The need to develop a reference capability for pictographs has led to many schemes and manipulations thereof based on some sort of accounting or atom-bonding balance sheet.

One family of such schemes includes those which consider molecules as being connected collections of "readily-recognizable" local assemblies of atoms such as chains,  $n$ -membered rings, functional groups (carbonyl, carboxylic acid, etc.), and the like. To such local assemblies are assigned characteristic symbols (digits, letters, or marks) so that the complete molecule is then simply represented by a collection of such symbols with suitable additional symbolism to describe their interconnections. To this particular family belong all the linear cypher notations such as those of Wiswesser,<sup>1</sup> Dyson,<sup>2</sup> IUPAC,<sup>3</sup> Bouman,<sup>4</sup> Gruber,<sup>5</sup> Gordon, *et al.*,<sup>6</sup> Silk,<sup>7</sup> and the like. The advantages of such systems include their compactness, ready correlation with molecular structure and current naming systems, and apparent ease of fragment searching or handling of partial indeterminates. On the debit side are the requirements for services of a trained intermediary for "molecular dissection," multiplicity of rules as to seniority or hierarchy of presentation and local conventions, necessity for a glossary or dictionary to define symbols, use of arbitrary nonchemically significant symbols and marks which are not meaningful or recognizable in all languages and which may conflict with standard chemical symbols, non-universality in that

only currently known structures have defined rules and new types of structures require rules supplementation, searching difficulties (ameliorated only partially by the permuted line entries concept<sup>8</sup>), and lack of good ordering conventions for a list of notations for different molecules. Another drawback is the lack of compatibility with the most widely used, chemically significant, and fundamental indexing parameter even for compounds of unknown structure, the molecular formula (although a suggested means of deriving this from the subject notations is implicit in a recent thesis and publication by Garfield<sup>9</sup>).

It must be recognized that all of the above schemes are simply extrapolations of the usual verbalizable chemical nomenclature with essentially all of the advantages and limitations thereof. It is good for relatively rapid communication between chemists knowledgeable in the rules, but fails when language barriers must be crossed or when one universally clear, unique, and acceptable name is to be placed in an ordered file.

Alternate and, possibly, increasingly more popular methods of describing molecular pictographs are those based on some sort of topological mapping of the individual atoms and the bonds between them. Into this category fall the "accounting" techniques of Ballard and Neeland,<sup>10</sup> and Gluck and Rasmussen,<sup>11</sup> the Polish tree-based method of Hiz,<sup>12</sup> and the matrix-based concepts of Ray and Kirsch,<sup>13</sup> and Sussenguth.<sup>14</sup>

It appears likely that such a representation based only on atoms and interconnecting bonds is the simplest mathematical transformation of a chemical pictograph. By "simplest" is meant that a minimal number of rules is required, essentially no more additional symbols than the usual chemical ones are utilized, and a clear-cut one-to-one correspondence with the chemical pictograph is obtained.

The present paper is devoted to a concept which belongs to the latter class of representations. It is initially based on a matrix construct which is then converted by standard algebraic techniques to a derivative form which has

\* Portions of this paper were presented before the Division of Chemical Literature, 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 11, 1963, and before the Chemical Information and Data System (CIDS) Phase IV Workshop, U. S. Army Missile Command, Huntsville, Ala., Nov. 12-14, 1963.

(1) W. J. Wiswesser, "A Line Formula Chemical Notation," T. Y. Crowell Co., New York, N. Y., 1954.

(2) G. M. Dyson, "A New Notation and Enumeration System for Organic Compounds," 2nd Ed., Longmans, Green and Co., New York, N. Y., 1958.

(3) International Union of Pure and Applied Chemistry, "Rules for I. U. P. A. C. Notation for Organic Compounds," John Wiley and Sons, Inc., New York, N. Y., 1961.

(4) H. Bouman, *J. Chem. Doc.*, **3**, 92 (1963).

(5) W. Gruber, *Angew. Chem.*, **61**, 429 (1950); "Die Genfer Nomenklatur in Chiffren," *Beihfte Angew. Chem. Chem.-Ind. Tech.*, No. **58**, 1950.

(6) M. Gordon, C. E. Kendall, and W. H. T. Davidson, "Chemical Ciphering," Royal Institute of Chemistry Monograph, 1948.

(7) J. A. Silk, *J. Chem. Doc.*, **3**, 189 (1963).

(8) P. F. Sorter, C. E. Granito, J. C. Gilmer, A. Gelberg, and E. A. Metcalf, *ibid.*, **4**, 56 (1964); 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 8-13, 1963, Abstracts 12G.

(9) E. Garfield, "An Algorithm for Translating Chemical Names to Molecular Formulas," Institute for Scientific Information, Philadelphia, Pa., 1961.

(10) D. L. Ballard and F. Neeland, *J. Chem. Doc.*, **3**, 196 (1963); *Chem. Eng. News*, **41**, 129 (April 15, 1963); 144th National Meeting of the American Chemical Society, Los Angeles, Calif., March 31-April 5, 1963, Abstracts 9F.

(11) *Chem. Eng. News*, **41**, 35 (Dec. 9, 1963).

(12) H. Hiz, *J. Chem. Doc.*, **4**, 173 (1964).

(13) L. C. Ray and R. A. Kirsch, *Science*, **126**, 814 (1957).

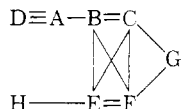
(14) E. H. Sussenguth, Jr., U. S. Army-Sponsored Conference on Chemical Information and Data System (CIDS), U. S. Army Missile Center, Huntsville, Ala., Nov. 12-14, 1963; 147th National Meeting of the American Chemical Society, Philadelphia, Pa., April 5-10, 1964, Abstracts 3F.

interesting and useful properties. A preliminary communication on this subject matter has appeared elsewhere.<sup>15</sup>

**The ACM.**—The core of this new symbolism is the Atom Connectivity Matrix, or ACM. This is a matrix array composed of mathematical elements<sup>16</sup>  $a_{ij}$ , where  $i$  and  $j$  are integers defining the row and column, respectively, at whose intersection  $a_{ij}$  is situated. In the diagonal elements, where  $i = j$ , are placed symbols for the atoms, radicals, electrons, charges, or other constituent groups making up the molecular formula. Order of assignment is unimportant. In the respective off-diagonal positions,  $a_{ij}$ , where  $i \neq j$ , are placed the "connectivity" (bond order, energy, force constant, directional cosines, or any other bond parameter) between the atom at  $a_{ij}$  and that at  $a_{ji}$ .

In most cases, bond order values of 0, 1, 2, and 3, to describe classical zero, single, double, and triple bonds, respectively, between two atoms or groups have proven to be the most useful "connectivity" parameters.

To describe the assembling of the ACM array, let us take a hypothetical complex "molecule"



where the connectivities pictured are conventional bond orders.

The atoms are placed into the diagonal elements in any order (the rows and columns are ruled for clarity).

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| A |   |   |   |   |   |   |   |
|   | B |   |   |   |   |   |   |
|   |   | C |   |   |   |   |   |
|   |   |   | D |   |   |   |   |
|   |   |   |   | E |   |   |   |
|   |   |   |   |   | F |   |   |
|   |   |   |   |   |   | G |   |
|   |   |   |   |   |   |   | H |

The appropriate connectivities, taken from the pictograph, are placed at the intersections of the rows and columns for the respective atoms. For example, the triple bond between A and D appears as

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| A |   |   | 3 |   |   |   |   |
|   | B |   |   |   |   |   |   |
|   |   | C |   |   |   |   |   |
| 3 |   |   | D |   |   |   |   |
|   |   |   |   | E |   |   |   |
|   |   |   |   |   | F |   |   |
|   |   |   |   |   |   | G |   |
|   |   |   |   |   |   |   | H |

since the 3 (for the triple bond) appears at the cross points for row and column 1 (governed by A) with row and column 7 (governed by the seventh atom, D). The complex interplay between atoms B, C, E, and F, where double bonds appear between B and C and between E and F, while single bonds occur between B and E, between B and F, between C and E, and between C and F, can be entered into the ACM as follows (note: the previous triple bond entry is not shown in the partial ACM in order to prevent confusion).

(15) L. Spialter, *J. Am. Chem. Soc.*, **85**, 2012 (1963).

(16) Throughout this paper, the term "element" refers to the mathematical concept of the matrix unit or entry, and not to its chemical meaning.

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| A |   |   |   |   |   |   |   |
|   | B | 2 |   | 1 | 1 |   |   |
|   | 2 | C |   | 1 | 1 |   |   |
|   |   |   | D |   |   |   |   |
|   | 1 | 1 |   | E | 2 |   |   |
|   | 1 | 1 |   | 2 | F |   |   |
|   |   |   |   |   |   | G |   |
|   |   |   |   |   |   |   | H |

The three-membered ring involving atoms C, F, and G would appear alone as

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| A |   |   |   |   |   |   |   |
|   | B |   |   |   |   |   |   |
|   |   | C |   |   | 1 | 1 |   |
|   |   |   | D |   |   |   |   |
|   |   |   |   | E |   |   |   |
|   |   | 1 |   |   | F | 1 |   |
|   |   | 1 |   |   | 1 | G |   |
|   |   |   |   |   |   |   | H |

The other atoms, A, B, D, E, and H, are not involved in this ring, as is reflected by the absence of any entries in their corresponding rows and columns.

The total ACM assembled for the hypothetical molecule in accordance with the principles exemplified above can then be written

|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| A | 1 |   | 3 |   |   |   |   |
| 1 | B | 2 |   | 1 | 1 |   |   |
|   | 2 | C |   | 1 | 1 | 1 |   |
| 3 |   |   | D |   |   |   |   |
|   | 1 | 1 |   | E | 2 |   | 1 |
|   | 1 | 1 |   | 2 | F | 1 |   |
|   |   | 1 |   |   | 1 | G |   |
|   |   |   |   |   | 1 |   | H |

Zeros can be placed in the empty spaces which indicate no bonds between the pertinent atoms. The resulting matrix is symmetric about the diagonal line of the atoms, and each half of the matrix contains all the connectivity of the molecular system. For computer input, it is sufficient then to cite only the main diagonal and the nonzero elements to one side of it.

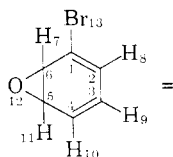
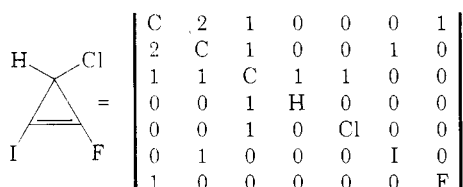
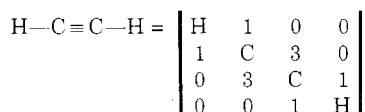
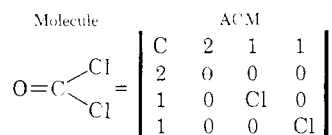
In the ACM, there are no numerical values in the diagonal because an atom is not considered as being bonded to itself. The off-diagonal elements of the ACM together with zero diagonal elements represent a typical connectivity matrix associated with a multiple graph (s-graph).<sup>17</sup> However, the special feature of the ACM is that the symbols identifying the atoms at the respective vertices of the pictograph are inserted and carried in the diagonal elements. The resulting ACM is a complete compact representation of the pictograph. Of course, the data can be further condensed by the scheme of Ray and Kirsch,<sup>13</sup> but this is a digression from development of the ACM concept.

The ACM is defined independently of atom-ordering priorities, numbering conventions, language, etc. It describes a molecular system, as does the usual chemical pictograph, in self-consistent internal coordinates, the constituent atoms or components and the interrelationships between them. Every atom is identified by its connectivities with all other atoms, and the final array, properly detailed, can represent only one skeletal linkage

(17) C. Berge, "The Theory of Graphs and Its Application," John Wiley and Sons, Inc., New York, N. Y., 1962, p. 130.

or bonded arrangement of its components.<sup>18</sup> The degree of detail is that determined by the original molecular pictograph and the ACM derived therefrom.

As illustrative examples of ACM's we may cite the following for the given molecular structures (where  $\theta$  is used instead of O for oxygen in the ACM in order to eliminate confusion with zero).



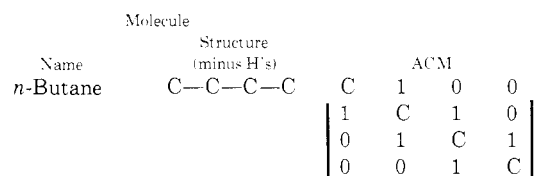
|    | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|
| 1  | C | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0  | 0  | 0  | 1  |
| 2  | 2 | C | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0  | 0  | 0  | 0  |
| 3  | 0 | 1 | C | 2 | 0 | 0 | 0 | 0 | 1 | 0  | 0  | 0  | 0  |
| 4  | 0 | 0 | 2 | C | 1 | 0 | 0 | 0 | 0 | 1  | 0  | 0  | 0  |
| 5  | 0 | 0 | 0 | 1 | C | 1 | 0 | 0 | 0 | 0  | 1  | 1  | 0  |
| 6  | 1 | 0 | 0 | 0 | 1 | C | 1 | 0 | 0 | 0  | 0  | 1  | 0  |
| 7  | 0 | 0 | 0 | 0 | 0 | 1 | H | 0 | 0 | 0  | 0  | 0  | 0  |
| 8  | 0 | 1 | 0 | 0 | 0 | 0 | 0 | H | 0 | 0  | 0  | 0  | 0  |
| 9  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | H | 0  | 0  | 0  | 0  |
| 10 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | H  | 0  | 0  | 0  |
| 11 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0  | H  | 0  | 0  |
| 12 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0  | 0  | H  | 0  |
| 13 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0  | 0  | 0  | Br |

(Numerals in the pictograph and those showing row and column are for clarity only. They are not included within the ACM.)

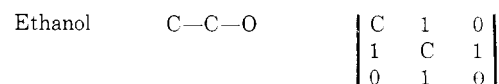
It is to be noted that the ACM for large molecules may appear as a formidable array. Nevertheless, this is not as bad as first glance might indicate because most of the off-diagonal elements are zero owing to the fact that, in a given row or column, the sum of the numerical values of all such elements must equal the total connectivity (or valency) of the atom in the diagonal element. Commonly, this is a small number not exceeding a value of 4.

However, it is possible to effect a significant saving in the size of the ACM by noting that its greatest application will probably be to organic molecules, both by

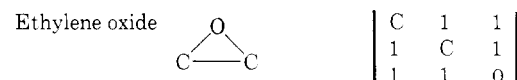
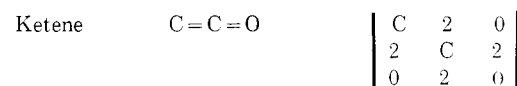
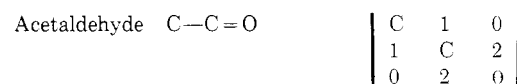
virtue of their overwhelming volume in the literature as well as their large numbers of isomers. Since hydrogen atoms, except in less common molecules such as perhalocarbons and others of similar nature, constitute a large proportion of the atoms in most organic structures (more than two-thirds in the case of the paraffinic hydrocarbons) and since their bond-order connectivity has a value of unity (*i.e.*, such atoms are not involved in the molecular backbone), it has been found possible and convenient to omit them from the ACM with no significant loss in generality for a majority of applications. The saving in time and size is evident from the following examples of "hydrogen-suppressed" ACM's.



(instead of a 14th-order ACM for C<sub>4</sub>H<sub>10</sub>.)



(instead of 9th-order ACM for C<sub>2</sub>H<sub>5</sub>OH)



Hydrogen suppression in pictographs is frequently used by chemists. In fact, almost all aromatic and many related ring compounds are generally written in this mode.

**ACM Applications to Other Molecular Types.**—In principle, any molecular pictograph and all of its accompanying data relating to atoms and bonds can be stored in the ACM. The ACM is an equivalent mathematical analog and the same questions may be asked of it as are asked of the pictograph. In essence, how the chemist draws the original molecule, its atoms, and their connectivities, determines also the ACM and its relative order of approximation to the molecule.

These concepts can probably best be explained by selecting specific applications to other molecular types. The same symbols will be used in the ACM as are conventionally used in the pictograph. Additional ones which may be found useful include  $\epsilon$ , for a localized electron (as in free radicals), and  $\text{ch}^{-n}$  and  $\text{ch}^{+n}$  to indicate localized charges of  $n$  electrical units. For specialized cases of stereoisomerism, imaginary bond orders may be used, as detailed in a separate communication on the subject.<sup>18</sup> It is worth noting that although special arbitrary symbolism

(18) The matter of stereoisomers and their treatment within the framework of the ACM concept will appear in a separate paper. For the present, spatial relations are not considered.

(19) The two-letter symbol  $\text{ch}$ , for charge, is a little awkward but it has mnemonic value. The Russian letter "cha" which has the "ch" pronunciation, might be better.

is here being introduced, these are utilized only for less commonly described molecular species, and may be modified or omitted as indicated.

Examples will be devoted to simple structure indexing. Since the rules are so simple and self-evident, discussion will be kept to a minimum.

#### 1. Free radicals

$$R\cdot = \begin{vmatrix} R & 1 \\ 1 & \epsilon \end{vmatrix} \quad \text{or, possibly better,} \quad |R\cdot|$$

#### 2. Ions

$$A^{-3} = \begin{vmatrix} A & 1 \\ 1 & ch^{-3} \end{vmatrix} \quad \text{or, possibly better,} \quad |A^{-3}|$$

#### 3. Covalent alkyl halides

$$R-X = \begin{vmatrix} R & 1 \\ 1 & X \end{vmatrix}$$

#### 4. Ionic salts

$$(A^{-})(B^{-}) = \begin{vmatrix} A & 0 & 1 & 0 \\ 0 & B & 0 & 1 \\ 1 & 0 & ch^{-1} & 0 \\ 0 & 1 & 0 & ch^{-1} \end{vmatrix}$$

This convention shows a zero-order bond between  $A^{-}$  and  $B^{-}$ . Whether or not to indicate the electrostatic bond between the charges  $ch^{-1}$  and  $ch^{-1}$  by some sort of symbol is a moot point. Of course, the pictograph above does not show one. However, inclusion of all these species in one ACM demonstrates that they comprise one compound. If localization of the charge element in the diagonal is not possible, then its off-diagonal elements probably should be made zero.

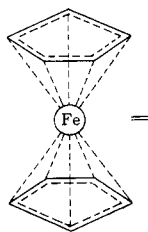
Alternatively, this species might be shown as

$$\begin{vmatrix} (A^{-}) & 0 \\ 0 & (B^{-}) \end{vmatrix} \quad \text{or} \quad \begin{vmatrix} (A^{-}) & 1 \\ 1 & (B^{-}) \end{vmatrix}$$

5. Clathrates, inclusion compounds, and linked ring chains wherein no specific bonds can be drawn between two separate molecular entities mechanically joined may be combined into one ACM with zero off-diagonal elements between them.

$$\textcircled{A}\textcircled{B} = \begin{vmatrix} (ACM)_A & 0 \\ 0 & (ACM)_B \end{vmatrix}$$

where  $(ACM)_A$  and  $(ACM)_B$  are the ACM's for components A and B, respectively. Conceptually, this is physically realistic since the diagonal elements contain all of the molecular components of the final compound and although rings A and B in the example are not joined by localized bonds, their integrity of structure are required for molecular stability. The ACM subfactors substantiate this idea.



|      |      |      |      |      |      |      |      |      |      |      |      |
|------|------|------|------|------|------|------|------|------|------|------|------|
| Fe   | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 | 1/10 |
| 1/10 | C    | 3/2  | 0    | 0    | 3/2  | 0    | 0    | 0    | 0    | 0    | 0    |
| 1/10 | 3/2  | C    | 3/2  | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| 1/10 | 0    | 3/2  | C    | 3/2  | 0    | 0    | 0    | 0    | 0    | 0    | 0    |
| 1/10 | 0    | 0    | 3/2  | C    | 3/2  | 0    | 0    | 0    | 0    | 0    | 0    |
| 1/10 | 3/2  | 0    | 0    | 3/2  | C    | 0    | 0    | 0    | 0    | 0    | 0    |
| 1/10 | 0    | 0    | 0    | 0    | 0    | C    | 3/2  | 0    | 0    | 3/2  | 0    |
| 1/10 | 0    | 0    | 0    | 0    | 0    | 0    | 3/2  | C    | 3/2  | 0    | 0    |
| 1/10 | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 3/2  | C    | 3/2  | 0    |
| 1/10 | 0    | 0    | 0    | 0    | 0    | 0    | 3/2  | 0    | 0    | 3/2  | C    |

Figure 2.

6. "Sandwich" compounds (ferrocene, dibenzenechromium, and the like), resonance hybrids, and similar delocalized species, where, however, connectivities can be specified even though they may be of fractional orders, can be transformed into their ACM's by the here-described technique [see, *e.g.*, the ACM for ferrocene (Fig. 2)]. It is here assumed for ferrocene that five delocalized electrons are distributed around each ring and that a sixth electron is delocalized between the iron and the five carbons of a ring. Since a bonding electron corresponds to one-half of a bond order unit, the C—C bonds in the ring have a bond order of 3/2, while the Fe—C is 1/10. Also, in accordance with the earlier mentioned approximation, the nonconnecting hydrogen atoms have been omitted from the pictograph and from the ACM. As this example shows, the pictograph model selected determines the ACM.

The extension to benzene in its delocalized form is exemplified in Fig. 1.

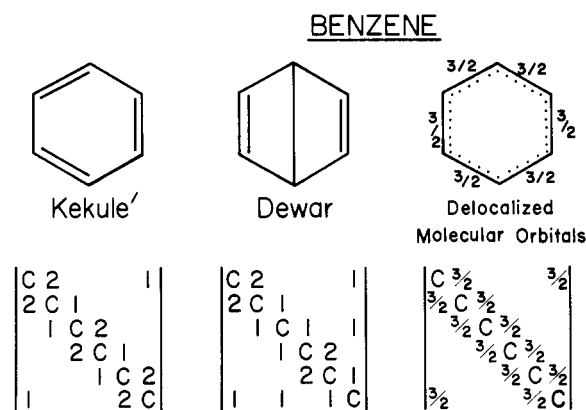
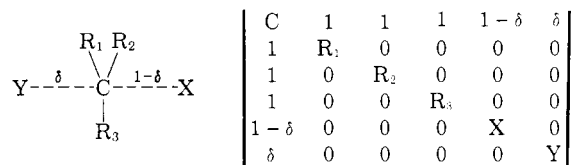
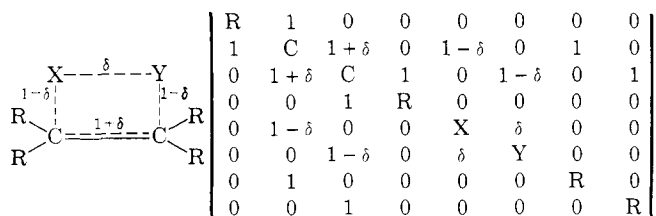


Figure 1.

7. Transition states, by virtue of their not being stable molecular species, require the use of a continuously changing connectivity parameter,  $\delta$ , which varies from 0 to 1 during the course of the reaction. A typical example is the transition state for the  $S_N2$  reaction.



For a four-centered transition state for elimination ( $E_1$ ), the following is obtained.



8. Macromolecules whose pictographs can be drawn completely can be transformed into their ACM's just as ordinary molecules are. Polymers of the  $n$ -mer type yield ACM's by taking the  $n$ th power of the ACM for the basic building block or monomeric unit. This latter technique is, of course, only an approximation since, in effect, it is assuming zero-order or nonlocalized bonding between the monomeric units. Schematic representation as  $|\text{ACM}|^n$  is sufficient.

The above examples are merely suggestions of applications. Certainly they are open to modifications and improvements which might have a salutary effect both on pictograph drawing and ACM conventions.

**Advantages and Defects of Storing Matrix Representations.**—A matrix notation for a molecule has certain excellent storageability characteristics which include the following:

(a) Simplicity of conversion to and from the pictograph. (The uniqueness of the one-to-one correspondence between the graph and the matrix appears to be well accepted.)

(b) Data stored consists of only numbers and a few symbols.

(c) Practically no computer time is required to convert input data to storage.

However, the matrix notation suffers from some very serious defects. Among these are the following:

(a) A group of matrices cannot be universally ordered; that is, beyond a preliminary screening by molecular formula, no method has been devised for organizing a collection of matrices in some numerical sequence defined by the matrices themselves. The difficulty is closely analogous to the unsolved problem of determining what pictograph is "closest" or "more like" a given pictograph. Chemically, it is similar to the question of finding the molecule "closest" to benzene. (Is it cyclohexane, pyridine, deuteriobenzene, hexadeuteriobenzene, toluene, cyclopentadienyl anion, fluorobenzene, 1,2-cyclohexadiene, or something else? Even more detailed specification of what is meant by "closeness to benzene," if short of specifying an exact molecule, does not resolve the ambiguity.)

The consequence of this lack of ordering is the great difficulty it places in the path of any search procedure. If one attempts to find whether or not a given matrix is in a collection, all those in the collection must be examined until *all* have been searched (if the matrix desired is actually not present) or until the desired one is found (if it is present). This is to be contrasted with searching in an internally ordered sequence, such as finding whether the number,  $X$ , is present in a sequence of numbers ordered by magnitude. One need look only at the point where all numbers smaller than  $X$  end and where all those larger begin. Whether  $X$  is present or absent is immediately discernible by examination at this point.

(b) The identity of the molecule represented by the pictograph is independent of how one cares to number or name the atoms in the molecule. This is not the case with a matrix type of representation. Order is implicit in the sequence in which the matrix, tree, or similar equivalent is assembled. The molecule, fluorochlorobromomethane, with five different atoms, can lead to 5! or 120 different-looking ACM's depending upon which permutation in the order of the constituent atoms is taken to fill the diagonal elements. Even the simple molecule, carbon disulfide, yields three apparently dissimilar ACM's.

$$\begin{array}{c}
 \left| \begin{array}{ccc}
 \text{C} & 2 & 2 \\
 2 & \text{S} & 0 \\
 2 & 0 & \text{S}
 \end{array} \right|
 \quad
 \left| \begin{array}{ccc}
 \text{S} & 2 & 0 \\
 2 & \text{C} & 2 \\
 0 & 2 & \text{S}
 \end{array} \right|
 \quad
 \left| \begin{array}{ccc}
 \text{S} & 0 & 2 \\
 0 & \text{S} & 2 \\
 2 & 2 & \text{C}
 \end{array} \right|
 \end{array}$$

The virtue of the ACM concept lies in the fact that each of these various permutations can be converted to any of the others for the same molecule by a paired number of row and column exchanges. This is a consequence of restricting the atoms in the matrix to the principal diagonal. Theoretically, such permutations could allow conversion of any matrix to a predetermined standard order in the elements and their bonding. For example, one might decide to assemble the atoms in the diagonal according to atomic number, and then organize similar atoms in such a sequence that the off-diagonal elements are clustered as close to the diagonal as possible. However, the necessity for trying all necessary permutations, with the additional cross-interactions with all other atoms (particularly in cases where so many atoms may be alike, as carbon in organics), makes a standard ordering scheme an exceedingly laborious task even if it is possible.

(c) Although only a few computer operations are required to store matrix data in any kind of memory device, very complex computer operations are required to search and compare matrices. Although a search routine of this type has been developed by Sussenguth,<sup>14</sup> it still represents a formidable and decreasingly efficient method as memory size increases because of both types of defects cited above.

(d) Another serious handicap of matrix notation, and one which is possibly insurmountable, is a consequence of the quantitative exactness introduced by the mathematical appurtenances thereto. These may be exemplified by reference to Fig. 1, which shows ACM's for various hydrogen-suppressed benzene pictographs. It is evident that the ACM's faithfully portray the quantitative significance of the pictograph but do not, and cannot, indicate that the chemist is referring to the same molecule in each case. In essence, unlike the "dissection" notation earlier mentioned where all of the graphs in Fig. 1 might be identified as benzenoid, the ACM implies only what the chemist draws, and it is assumed that the chemist means what he draws and draws what he means. A given connectivity has precise mathematical meaning and the ACM does not uncover new chemical theories or verities. For example, resonance, keto-enol tautomerism, clathrate and coordinate structures,  $\text{S}_\text{N}2'$  reactions, and other such theoretical concepts are not revealed to the chemist (although the ACM has some superficial similarity to matrix constructs in LCAO molecular orbitals).

A concomitant of the quantitative mathematical meaning to the ACM appears when one attempts to treat polynuclear aromatics such as naphthalene and phenan-

threne. One has difficulty in deciding proper values of connectivities to adopt. The use of Kekulé-type structures requires a matrix for every canonical form and is not in accord with experimental bond orders. On the other hand, experimental bond orders may vary, albeit slightly but significantly, from time to time and change matrix entries so that such experimental connectivities are not suitable for a storage-search-retrieval system. On the other hand, an arbitrary fixed value, such as  $\frac{3}{2}$  for an aromatic bond connectivity, may obscure the fact that the 9,10-bond in phenanthrene and the 1,2-bond in naphthalene frequently behave more like double bonds than like aromatic ones. This may require cross-referencing such cases into a limited number of related ACM's to represent alternate acceptable "equivalent" pictographs.

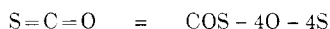
It is evident that the storage simplicity of the matrix formulation carries with it certain serious limitations with respect to search and retrieval operations. Some scheme with the advantages but not the defects of the matrix method is desirable. One such concept, here proposed, appears to meet these requirements except for overcoming the last cited difficulty above. It is a mathematical "relative" to the ACM.

**The ACMCP.**—The one quantity derivable from the ACM, considered as a determinant, which is invariant (as is the chemical pictograph) with respect to permutations in the ordering of the constituent atoms, is the characteristic or determinantal polynomial. For convenience we have dubbed it, in the present application, the Atom Connectivity Matrix Characteristic Polynomial (ACMCP). This polynomial is obtained by evaluating the matrix as a determinant through the usual mathematical techniques, such as decomposition into minors and the summation of factor-cofactor products,<sup>20</sup> or by matrix operations not requiring direct expansion.<sup>21</sup> In matrix calculus it is known as the "determinantal polynomial" or "value of the determinant" of the matrix. Symbolically

$$\text{ACMCP} = |\text{ACM}| = \sum_{p(i)} (-1)^v a_{1i_1} \cdot a_{2i_2} \cdot \dots \cdot a_{ni_n}$$

where ACMCP is the Atom Connectivity Matrix Characteristic Polynomial;  $|\text{ACM}|$  means the value of the determinant of the ACM matrix;  $a_{ni}$  is the element in row  $n$ , column  $i$ , of the ACM;  $i_1, i_2, \dots, i_n$  are the column subscripts 1, 2,  $\dots$ ,  $n$  taken in some order, *i.e.*, a permutation;  $v$  is the number of inversions in the above permutation; and the indicated algebraic sum is taken over all possible permutations of the subscripts. This range of summation is denoted by  $p(i)$ .

In the cited example of carbonyl sulfide, we may equate the molecular pictograph to the analogous ACMCP determined in accordance with the above definition



This same ACMCP is obtained from any of the ACM's constructed for this molecule.

Other examples of ACMCP's for complete simple molecules are shown in Table I. The initial ACM's for the

Table I  
ACMCP's for Complete Molecules

| Molecule        | Pictograph   | ACMCP   |
|-----------------|--|---|
| Methane         | $\begin{array}{c} \text{H} \\   \\ \text{H}-\text{C}-\text{H} \\   \\ \text{H} \end{array}$  | $\text{CH}^4 - 4\text{H}^3$   |
| Acetylene       | $\text{H}-\text{C}\equiv\text{C}-\text{H}$   | $\text{C}^2\text{H}^2 - 2\text{CH} - 9\text{H}^2 + 1$   |
| Formaldehyde    | $\text{H}_2\text{C}=\text{O}$  | $\text{CH}^2\text{O} - 2\text{HO} - 4\text{H}^2$  |
| Methyl chloride | $\begin{array}{c} \text{H} \\   \\ \text{H}-\text{C}-\text{Cl} \\   \\ \text{H} \end{array}$ | $\text{CH}^3\text{Cl} - \text{H}^3 - 3\text{H}^2\text{Cl}$                                    |
| Cyanogen        | $\text{N}\equiv\text{C}-\text{C}\equiv\text{N}$  | $\text{C}^2\text{N}^2 - 18\text{CN} - \text{N}^2 + 81$  |
| Ketene          | $\text{H}_2\text{C}=\text{C}=\text{O}$   | $\text{C}^2\text{H}^2\text{O} - 4\text{CH}^2 - 4\text{H}^2\text{O} - 2\text{CHO} + 8\text{H}$ |

molecules listed therein were prepared in accordance with the earlier described rules.<sup>13</sup>

It is to be noted that the first term in the ACMCP is always equivalent to the molecular formula of the compound (in the French style where superscripts are used instead of subscripts) since it is the product of all the atoms or groups in the diagonal of the ACM.<sup>22</sup> Hence, molecular formula indexes represent tabulations of lowest order approximations to the pertinent ACMCP's. Further, because of the unidirectional one-to-one mapping into the ACMCP from its molecular progenitor, it is conceivable that a listing of ACMCP's with literature data and abstract citations thereunder could, with one tabulation, replace both the separate formula index and the compound name index commonly used, as far as a large memory and storage facility is concerned.<sup>23</sup> The former index is already overburdened under most entries with a multiplicity of named isomers frequently requiring a pictograph to be drawn for clear comprehension even of nonpertinent citations; the latter index suffers from frequent name changes and arbitrary numbering of positions, the complications introduced by the ever-growing list of trivial names, the clashing and confusing rules of nomenclature, language, ordering and priority, and the incompatibility with efficient information storage and retrieval techniques.

**Hydrogen-Suppressed ACMCP (HS-ACMCP).**—The ACM for any molecule (or reaction) can be readily written, even for large molecules, because of the simple rules involved and because the limited valency or bonding capacity of any atoms results in an increasing proportion of off-diagonal zeros with increasing matrix order (*i.e.*, molecular size). However, the ACMCP, although its coefficients are generally integers (particularly for classical structures) and the number of terms of different exponential degree never exceeds the number of atoms in the

(20) (a) C. C. MacDuffee, "The Theory of Matrices," Chelsea Publishing Co., New York, N. Y., 1956, p. 17 ff.; (b) W. V. Parker and J. C. Eaves, "Matrices," The Ronald Press Co., New York, N. Y., 1960, p. 122 ff.

(21) (a) H. E. Fettis, *Quart. Appl. Math.*, **8**, 206 (1950); (b) J. S. Frame, *Bull. Am. Math. Soc.*, **55**, 1045 (1949).

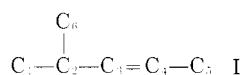
(22) The order of writing the atomic symbols in each term of the ACMCP and the order of writing the terms of a given degree may, of course, follow any desired convention, such as alphabetical order, increasing atomic number, or frequency of usage. Some arbitrary standard will have to be adopted.

(23) In a private communication, Dr. Fred A. Tate, Assistant Director, Chemical Abstracts Service, has estimated that 80-85% of the chemical literature can be indexed under names of chemical compounds.

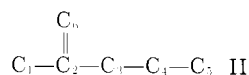
diagonal elements,<sup>24</sup> may possess for a given exponential degree a variety of terms. Note, for example, the ACMCP in Table I for ketene, wherein the third degree terms are:  $-4\text{H}^2\text{O} - 4\text{CH}^2 - 2\text{CH}\text{O}$ . It is characteristic of the ACMCP, as will be shown later,<sup>25</sup> that the nonnumerical (*i.e.*, lettered) factors of terms beyond the first are derived chiefly from combinations of atoms which have few connectivities. Hence, monovalent or singly bonded atoms such as hydrogen and the halogens show up in profusion among these terms. Since hydrogen is not a molecular skeletal binding component, except in cases of hydrogen bonding, it can, as was mentioned earlier, be omitted from the ACM. This elimination diminishes the total number of terms and simplifies their nature in the ACMCP. However, as a result of dropping or "suppressing" the hydrogen atoms, some information is lost and must be compensated for in other ways. The informational loss consists of the following:

(a) The hydrogen count which entered into the first term, among others, and made this term identical with the molecular formula is no longer carried implicitly through the computation from the ACM. To maintain this useful attribute, it is necessary and expeditious to introduce externally into the first term the requisite number of hydrogens (as a factor in the symbol H, for hydrogen, raised to the appropriate power). This added factor, in powers of H, appears only in the first term in this modification to the "hydrogen-suppressed" ACMCP.

(b) A more fundamental and serious loss is that information concerning the total valency of each atom in a given molecule. For example, except for a few exceptions such as isonitriles, carbon monoxide, metal carbonyls, and radicals or molecular fragments, carbon exhibits a valence of four in organic compounds. When the hydrogen is suppressed, however, a pictograph may show carbon symbols, C, with apparent valences varying from one up to four. This may be illustrated by the structure (HS or hydrogen-suppressed) for 4-methyl-2-pentene (I)



wherein carbons numbered 1, 5, and 6 appear to be univalent, and carbons numbered 2, 3, and 4 are tervalent. The consequences of losing this valency data by treating all C symbols as equivalent in valency, despite the acknowledged falsity of this assumption, and multiplying them together to give terms such as  $\text{C}^6$ , is the development of serious ambiguities and degeneracies in the ACMCP. For example, HS 2-methyl-1-pentane (II)



yields the *same* HS-ACMCP as does its isomer (I):  $\text{C}^6 - 8\text{C}^4 + 11\text{C}^2$ . Other similar examples can be found as well

as some artificial mathematical constructs which are subject to accidental peculiarities in number theory, such as  $\text{X}-\text{X} \equiv \text{X}$  and  $\text{X}-\text{X} \equiv \text{X}$ , both of whose ACMCP's equal<sup>26</sup>  $\text{X}^3 - 50\text{X}$ , partly because  $1^2 + 7^2 = 5^2 + 5^2 = 50$ . Again, in this latter case, there is also the basic defect of allowing a given symbol, X, to vary its valency over values of 1, 5, 7, 8, and 10 in the two pictographs, yet considering the symbols to be algebraically, or computationally, equivalent.

Several ways to resolve these hydrogen-suppressed ACMCP degeneracies present themselves:

(a) One could ignore the ambiguities, since empirical checks suggest that in physically real molecules, where atoms possess only small apparent valency magnitudes (almost always less than 5) and these in only limited number, such cases occur only a small fraction of the time (possibly less than 0.1%). The duplications could then be sorted out at information retrieval time by manual examination. Such a crude solution is subject to error and is not philosophically or conceptually satisfying. It must be rejected.

(b) Certain of the degeneracies, but not all, can be broken by inserting the number of hydrogens suppressed into the first term. However, isomeric cases such as those typified by compounds I and II are not resolvable by this means.

(c) One could retain all valency information by not adopting the hydrogen-suppression technique but by including *all* atoms in the ACM from which the ACMCP is derived. While it is conjectured<sup>27</sup> that this practice will eliminate the cited problems, it, nevertheless, will result in exceedingly large and cumbersome ACMCP's. This is highly undesirable for both computational and storage reasons, although it may eventually prove to be necessary.

(d) Retaining consecutively numbered subscripts on atoms of the same symbol breaks the degeneracy but leads to ACMCP's which are both cumbersome and no longer invariant to numbering permutations.

(e) Another means of retaining the lost valency information is to assign different symbols for atoms of the same kind which, however, exhibit different apparent valences in the HS case. For example, in compound I,  $\text{C}_1$ ,  $\text{C}_5$ , and  $\text{C}_6$ , being univalent, can be called  $\text{C}(1)$ ; tervalent  $\text{C}_2$ ,  $\text{C}_3$ , and  $\text{C}_4$  can be called  $\text{C}(3)$ . The resulting HS-ACMCP is  $\text{C}(1)^3\text{C}(3)^3 - [3\text{C}(1)^2\text{C}(3)^2 + 5\text{C}(1)^3\text{C}(3)] + [9\text{C}(1)^2 + 2\text{C}(1)\text{C}(3)]$ . By an analogous convention, compound II yields as its HS-ACMCP,  $\text{C}(1)^3\text{C}(2)^3\text{C}(4) - [5\text{C}(1)^2\text{C}(2)^2 + \text{C}(1)\text{C}(2)^3 + \text{C}(1)^2\text{C}(2)\text{C}(4) + \text{C}(1)\text{C}(2)^2\text{C}(4)] + [4\text{C}(1)^2 + 6\text{C}(1)\text{C}(2) + \text{C}(2)^2]$ . Although this method can be used, the enlargement of the ACMCP in this manner is neither desirable nor convenient.

(f) A technique which appears to resolve all the ambiguities found to date, yet retains the simplicity of ordinary symbolism, is one which appends to the HS-ACMCP a

(24) This is readily proven by the fact that an  $n$ th order matrix can yield at most an  $n$ th degree polynomial containing  $n + 1$  coefficients when the terms are grouped by total exponential degree. However, since the ACM diagonal elements contain no numerical additive component other than the constituent atoms (groups or particles), the "trace" (sum of such numerical additive components) of the matrix is zero, and the coefficient of the  $(n - 1)$ th degree term (determined by the trace) in the ACMCP is also zero. Hence, the ACMCP contains only  $n$  differently ordered terms with the first term being of the  $n$ th degree and the next one of the  $(n - 2)$ th degree.

(25) L. Spialter, *J. Chem. Doc.*, 4, 269 (1964).

(26) This simple case was pointed out to the author by Dr. Sylvan Eisman, Pitman-Dunn Institute for Research, U. S. Army, Frankford Arsenal, Pa.

(27) No mathematical proof for the uniqueness of the transformation from the total ACMCP to the initial pictograph (where equivalent symbolism carries with it equivalent valency) has yet been obtained. However, since the ACMCP is invariant with respect to the defined permutations of the ACM (which leave only atomic symbols in the principal diagonal), the mapping from pictograph to ACMCP is single-valued for a given type of symbolism and quantitative standard for the pictured connectivities.

table identifying types and numbers of atoms of the different apparent valencies. For example, the one quadrivalent, no tervalent, three bivalent, and two univalent carbon atoms in compound II can be assembled into a "valency count" as  $C(1, 0, 3, 2)$ . The comparable valency count for compound I is  $C(0, 3, 0, 3)$ . By inclusion of this simple concept, *all* degeneracies observed to date in HS-ACMCP's have been broken. This artifice now appears to retain the virtues of convenience and reliability.<sup>28</sup>

It is now proposed that, for the hydrogen-suppressed cases, the appropriate polynomial to use as a computer-oriented name is the resultant ACMCP with the inclusion of hydrogen count in the first term and an appended valency count (or VC) collection. To this assemblage has been given the name "modified hydrogen-suppressed, or MHS-ACMCP. A list of some MHS-ACMCP's for various hexadecenes is presented in Table II. The conjecture is that if two pictographs have identical MHS-ACMCP's, they represent identical molecules at least to the extent to which information has been entered into the ACM,<sup>29</sup> which may include ignoring stereoisomeric data.<sup>18</sup>

It is interesting and useful to point out that, as can be seen from Table III, the polynomial portion of the

MHS-ACMCP can be derived from the complete ACMCP by dividing all terms beyond the first by H raised to the hydrogen count (exponent) and eliminating all those resulting terms which contain a negative power of H as one of their factors.

**Other ACMCP Approximations.**—There are two other simple ACMCP approximation schemes, which, while not uniquely sufficient in identifying a molecular compound or system, are necessary to be satisfied. As such, they are adaptable to small or personal data collecting and referral or to very rapid screening processes.

Particularly suitable for the latter purpose is the "condensed" ACMCP, wherein the first term (or molecular formula) is first written down and is then followed by the algebraic sum of the coefficients of the remaining terms. By this procedure, condensation of the appropriate complete ACMCP's gives  $C^2H^2O - 2$  for ketene, and  $C^5H^6 - 9$  for cyclopentadiene (refer to Table III). The modified hydrogen-suppressed ACMCP's yield, respectively,  $C^2H^2O - 8$  and  $C^5H^6 + 23$ , by compression. The two systems give differing results and, hence, operational conventions must be adopted.

Another method of approximation, equally adaptable to

Table II  
Modified Hydrogen-Suppressed ACMCP's for Various Isomeric Hexadecenes<sup>a</sup>

| Name                 | ACMCP Coefficients |       |       |       |       |       |       |       |       |       | VC            |
|----------------------|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---------------|
|                      | $C^2H^2$           | $C^1$ | $C^2$ | $C^3$ | $C^4$ | $C^5$ | $C^6$ | $C^7$ | $C^8$ | $C^9$ |               |
| 3-Propyl-1-tridecene | 1                  | -18   | +126  | -444  | +842  | -840  | +388  | -56   | 0     | 0     | $C(0,2,12,2)$ |
| 2-Pentyl-1-undecene  | 1                  | -18   | +126  | -444  | +842  | -844  | +404  | -68   | 0     | 0     | $C(1,0,13,2)$ |
| 2-Heptyl-1-nonene    | 1                  | -18   | +126  | -444  | +842  | -844  | +404  | -72   | 0     | 0     | $C(1,0,13,2)$ |
| 2-Hexyl-1-decene     | 1                  | -18   | +126  | -444  | +842  | -844  | +404  | -72   | +4    | 0     | $C(1,0,13,2)$ |
| 2-Butyl-1-dodecene   | 1                  | -18   | +126  | -444  | +842  | -844  | +404  | -80   | +4    | 0     | $C(1,0,13,2)$ |
| 2-Hexadecene         | 1                  | -18   | +127  | -451  | +855  | -840  | +378  | -57   | +1    | 0     | $C(0,2,12,2)$ |
| 4-Hexadecene         | 1                  | -18   | +127  | -454  | +879  | -903  | +438  | -72   | +1    | 0     | $C(0,2,12,2)$ |
| 6-Hexadecene         | 1                  | -18   | +127  | -454  | +879  | -906  | +450  | -81   | +1    | 0     | $C(0,2,12,2)$ |
| 8-Hexadecene         | 1                  | -18   | +127  | -454  | +879  | -906  | +450  | -84   | +1    | 0     | $C(0,2,12,2)$ |
| 7-Hexadecene         | 1                  | -18   | +127  | -454  | +879  | -906  | +450  | -84   | +4    | 0     | $C(0,2,12,2)$ |
| 5-Hexadecene         | 1                  | -18   | +127  | -454  | +879  | -906  | +453  | -90   | +4    | 0     | $C(0,2,12,2)$ |
| 3-Hexadecene         | 1                  | -18   | +127  | -454  | +882  | -924  | +483  | -102  | +4    | 0     | $C(0,2,12,2)$ |
| 1-Hexadecene         | 1                  | -18   | +130  | -484  | +990  | -1092 | +588  | -120  | +4    | 0     | $C(0,1,14,1)$ |

<sup>a</sup> *cis-trans* isomerism is ignored. <sup>b</sup> All terms of odd degree have zero coefficients. <sup>c</sup> Valency count (see text).

Table III  
Comparison of ACMCP's for Various Molecules

| Compound           | Complete ACMCP   | MHS-ACMCP                                    |
|--------------------|--|--|
| Ketene             | $C^2H^2O - (4CH^2 + 4H^2O + 2CHO) + 8H$  | $C^2H^2O - (4C + 4O); C(1,0,2,0)$            |
| 3-Iodocyclopropene | $C^3H^3I - (C^3H^3 + 3C^2H^3I + 6CH^3I) + 4H^3I + (2CH^2 + 3CHI + 4H^3 + 6H^2I) - (H + I)$   | $C^3H^3I - (C^2 + 6CI) + 4I + 4; C(0,3,0,0)$ |
| Cyclopentadiene    | $C^5H^6 - (6C^4H^5 + 11C^3H^6) + (14C^2H^4 + 42C^2H^5 + 26CH^6) + 8H^6 - (18C^3H^3 + 49CH^4 + 42H^5) + (9CH^2 + 20H^3) - 2H$                                   | $C^5H^6 - 11C^3 + 26C + 8; C(0,4,1,0)$       |
| Benzene (Kekulé)   | $C^6H^6 - (6C^5H^5 + 15C^4H^6) + (15C^4H^4 + 60C^3H^5 + 63C^3H^6) - (20C^2H^3 + 90C^2H^4 + 126CH^5 + 81H^6) + (15C^2H^2 + 60CH^3 + 63H^4) - (6CH + 15H^2) + 1$ | $C^6H^6 - 15C^4 + 63C^2 - 81; C(0,6,0,0)$    |

<sup>a</sup> Modified hydrogen-suppressed ACMCP (see text).

(28) Although only comparison of valency count for carbon was empirically found to be sufficient for this purpose, it is probably safest to list the valency count for all atoms wherein the "apparent valency" differs from the actual value in the molecule because of hydrogen suppression. Such VC data are readily available from the ACM because they involve simply a tabulation of the sum of the row (or column) connectivity entries for each atom in the diagonal.

(29) This concept suggests the direction in which to seek a mathematical proof for the uniqueness of the MHS-ACMCP. It should be necessary and sufficient to establish that, given the set of atoms or points defined by a given valency count, no two pictographs or linear graphs can be constructed therefrom which yield the same ACMCP. Assistance in this proof may be rendered by utilization of the physico-geometric significance of the ACMCP which is presented elsewhere.<sup>12</sup>



fast computer scanning, is the "truncated" ACMCP. This is obtained by considering only the first three of four terms of the ACMCP for indexing and then hand-sorting for further selection. This approximation method is particularly convenient because, with techniques described elsewhere,<sup>25</sup> it is possible to compute the value of the first few coefficients by simple inspection of the molecular pictograph without the necessity of constructing the ACM and reducing it to the polynomial. This visual computation technique is rapid, screens a large quantity of data efficiently, and yet maintains a close discernible rapport between the chemical and the mathematical concepts of the ACM-ACMCP nomenclature. As such it lends itself to personal filing systems, commercial chemical catalog listing, and the like.

It must be emphasized that most of the specific details of formulation and procedure discussed above need not be the concern of the average chemist. He is free to converse with his colleagues by means of molecular pictographs, as

he now does, even in abstracts,<sup>30</sup> or by compound names, systematic or trivial, but he leaves most of the literature searching and data retrieval—particularly that related to specific compounds—to computer-based systems. His link to and from the computer is the pictograph, but the ACM-ACMCP concept is suggested as providing the working language within the storage-retrieval system.

**Acknowledgments.**—Acknowledgment is gratefully made to the Applied Mathematics Laboratory, ARL, for assistance in the more sophisticated development of some material in this paper. In particular should be mentioned the computer programming help given by Henry E. Fettes and James C. Caslin, the location of many hydrogen-suppressed ACMCP redundancies by Lt. Donald L. Barnett, and the excellent sounding board qualities of Major John V. Armitage, all of the Laboratory.

(30) Note, for example, the publication "Index Chemicus," Institute for Scientific Information, Philadelphia, Pa.

## The Atom Connectivity Matrix Characteristic Polynomial (ACMCP) and Its Physico-Geometric (Topological) Significance\*

By LEONARD SPIALTER

Chemistry Research Laboratory, ARL, Wright-Patterson Air Force Base, Ohio

Received February 24, 1964

### INTRODUCTION

The concept of the Atom Connectivity Matrix (ACM) and its characteristic polynomial (ACMCP) as a computer-oriented nomenclature for chemical pictographs has been described elsewhere.<sup>1,2</sup> The ACM is an array in which symbols for the constituent atoms or vertices of a pictograph (chemical or otherwise) are placed as mathematical elements in the main diagonal, and values or symbols for the respective connectivities for pairs of atoms are entered at appropriate off-diagonal column-row intersections. The ACMCP is derivable from the ACM by any of a number of algebraic processes.<sup>3</sup>

The ACM is clearly directly relatable, datum-by-datum, to its pictographic progenitor. However, it appeared initially that, in the process of transforming the ACM into the ACMCP, any physical significance would be lost and the ACMCP would be simply a collection of mathematical artifacts, exceedingly useful but physically and chemically meaningless. Such proved *not* to be the case.

**The Homoatomic ACMCP.**—If one takes a homoatomic<sup>4</sup> ACM of completely general connectivity, such as the fifth-order symmetric ACM<sup>5</sup>

$$\begin{vmatrix} X & p & q & r & s \\ p & X & t & u & v \\ q & t & X & w & y \\ r & u & w & X & z \\ s & v & y & z & X \end{vmatrix}$$

(where X represents the homoatom; *i* and *j* are running indexes (from 1 to 5) to identify row and column, respectively; and *p, q, r, . . . , w, y, z* represent the appropriate connectivity between the *X<sub>i</sub>* and *X<sub>j</sub>* atoms) and determines the coefficients of each degree in X, in the related ACMCP, the following results are obtained.

$$\begin{aligned} \text{Coeff. of } X^5: & 1 \\ \text{Coeff. of } X^4: & 0 \\ \text{Coeff. of } X^3: & -(p^2 + q^2 + r^2 + s^2 + t^2 + u^2 + v^2 + w^2 + y^2 + z^2) \\ \text{Coeff. of } X^2: & +2(pqt + prv + psu + qrw + qsv + rsz + tuw + tvy + \\ & \quad uwz + wyz) \\ \text{Coeff. of } X: & -[2(pquw + pqvy + prtw + pruz + psty + psuz + \\ & \quad qrut + qryz + qstv + qsuw + rsuv + rsuw + tuyz + \\ & \quad twvz + uvwy) - \\ & \quad (p^2w^2 + p^2y^2 + p^2z^2 + q^2u^2 + q^2v^2 + q^2z^2 + r^2t^2 + \\ & \quad r^2v^2 + r^2y^2 + s^2t^2 + s^2u^2 + s^2w^2 + t^2z^2 + u^2y^2 + \\ & \quad v^2w^2)] \\ \text{Constant term:} & +2[(pqvyz + pqvuz + prtyz + prvuy + pstwz + \\ & \quad psuwy + qrtvz + qruvy + qstuz + qsuwv + rstuy + \\ & \quad rstvw) - (p^2wyz + q^2uwz + r^2tvz + s^2tuw + t^2rsz + \\ & \quad u^2qsy + v^2qrw + w^2psv + y^2pru + z^2pqt)] \end{aligned}$$

\* A portion of this subject material was presented before the Chemical Abstracts Research and Development Division, Columbus, Ohio, Aug. 15, 1963, and before the Chemical Information and Data System (CIDS) Phase IV Workshop, U. S. Army Missile Command, Huntsville, Ala., Nov. 12-14, 1963.

(1) L. Spialter, *J. Am. Chem. Soc.*, **85**, 2102 (1963).

(2) L. Spialter, *J. Chem. Doc.*, **4**, 261 (1964).

(3) Methods and mathematical nomenclature herein used are to be found in any text or reference on matrix algebra, matrix calculus, or determinants. For the sake of brevity and concentration on the newer concepts introduced in this work, standard operations will not be detailed.

(4) For the sake of simplicity in understanding the initial development, it is preferable that all atoms in the diagonal be made equivalent. Extension to heteroatomic systems will be presented later.

(5) The symmetric case is representative of a molecule or transition state.<sup>6</sup>