

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,²⁵ 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,³⁰ 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.^{1,2} 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.³

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 homoatomic⁴ ACM of completely general connectivity, such as the fifth-order symmetric ACM⁵

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

(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_i and X_j atoms) and determines the coefficients of each degree in X, in the related ACMCP, the following results are obtained.

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

* 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.^{1,2}

Now, since the above generalized ACM is equivalent to the schematic pictograph shown in Fig. 1,⁶ it is possible to assign physico-geometric significance to each coefficient in the ACMCP.

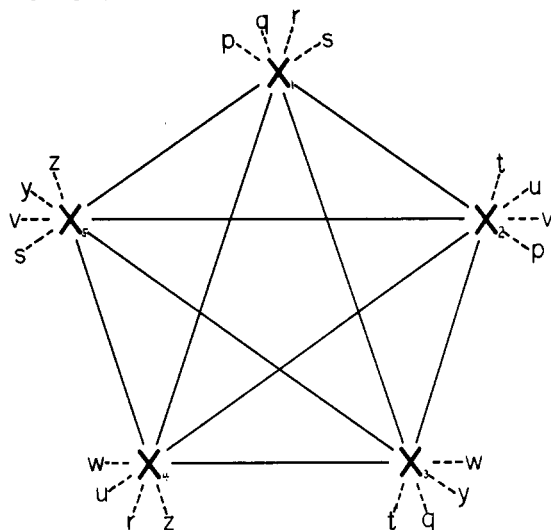


Fig. 1.—Schematic pictograph of generalized five-centered homoatomic interconnectivity model.

Since the highest degree term is derived from multiplication of all the elements of the principal diagonal, its coefficient is always unity for single atoms or groups filling the diagonal and having individual coefficients of unity (probably the exclusive foreseeable situation).

The coefficient of X^4 is zero because the numerical "trace" of the matrix is zero.³

The coefficient of X^3 is the negative of the sum of the squares of each of the bond connectivities. This can be computed directly from visual examination of the pictograph. If a connectivity has a value of zero, it represents a "no-bond" and its contribution is zero whether it is drawn into the pictograph with a zero value, or is omitted entirely.

The coefficient for the third term of the subject ACMCP, that in X^2 , is twice the sum of the bond connectivity products of each three-membered ring in the pictograph. This can be deduced by examining the significance of the three-factor products of this term on the pictograph of Fig. 1. For example, the term pqt represents the product of bond orders around a closed three-membered ring. If a molecular pictograph contains no such complete ring, then the coefficient for this term becomes zero.

Similarly, it can be seen from Fig. 1 that the next coefficient is the negative of twice the sum of the bond connectivity products around each four-membered ring (analogous to the three-membered ring case above), to which is then added the sum of the squared products of all nonadjacent bond connectivities taken two at a time. The adjective "nonadjacent" means that the pairs of bond connectivities do not have an atom in common. For example, p, v, u , and t are all adjacent connectivities since they share the common atom, X (numbered 2). Since the only bonds nonadjacent to p are w, y , and z , the subject

coefficient contains the terms, p^2w^2 , p^2y^2 , and p^2z^2 , as the only ones of this type involving p .

The coefficient for the last term, that is, the constant for the generalized fifth-order ACMCP, consists of two components: those like the expression $pquyz$, which represent the product of bond orders around a five-membered ring, and those like p^2wyz which represent the product of bond orders around a three-membered ring multiplied by the square of a nonadjacent connectivity.

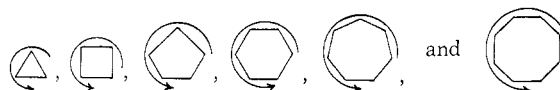
Although the various coefficients for the generalized homoatomic ACMCP for n -atoms can be written as above (or in standard mathematical nomenclature), it has been found more useful to develop a geometrical notation by extrapolation and generalization from the above-discussed concepts. This permits the chemist to bypass the ACM and to construct the coefficients of the ACMCP directly from the pictograph for the homoatomic case while retaining the concept of the classical geometrical molecular structural components, such as rings, multiple bonds, and the like. The heteroatomic case is analyzed later.

The values of the first nine coefficients, in the geometric notation, are shown in Table I. The symbol $(-)$ ² refers to

Table I
The Physico—Geometric ACMCP Coefficients

Degree in X	Coefficient
n	1
$n-1$	0
$n-2$	$-\sum (-)^2$
$n-3$	$+2\sum \triangle$
$n-4$	$-2\sum \square + \sum (-)_a^2 (-)_b^2$
$n-5$	$+2\sum \triangle - \sum (-)_a^2 \triangle_b$
$n-6$	$-2\sum \square + 2\sum (-)_a^2 \square_b$ $+4\sum \triangle_a \triangle_b - \sum (-)_a^2 (-)_b^2 (-)_c^2$
$n-7$	$+2\sum \triangle - 2\sum (-)_a^2 \triangle_b$ $-4\sum \triangle_a \triangle_b + 2\sum (-)_a^2 (-)_b^2 \triangle_c$
etc.	

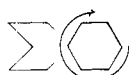
the square of a particular connectivity (bond order in these cases). By analogy with the symbols which follow, this symbol can also be considered as the product of connectivities around a "two-membered" ring. The geometric symbols



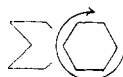
(6) The pentagonal pattern is simply one way of drawing the array. The structural invariant is the set of atoms, X , and their respectively lettered interconnectivities. In Fig. 1, the atoms are numbered to correspond with their arbitrarily placed order in the ACM's diagonal.

refer to the products of bond orders around the perimeter periphery of three-, four-, five-, six-, seven-, and eight-membered rings, respectively, in the molecule whose ACMCP is being computed. The subscripts a, b, c, d, indicate that the geometric groups concerned with each term are not adjacent to each other; *i.e.*, they are not attached to a common atom. The summation, symbolized by \sum , is taken over all acceptable sets of pictured relationships, without duplication, which are present in the molecule.

By way of example, the



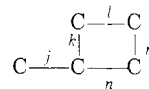
terms for cyclohexane, cyclohexene, 1,3-cyclohexadiene, 1,4-cyclohexadiene, and benzene (Kekulé form) are 1, 2, 4, 4, and 8, respectively, as obtained by multiplying the bond orders around the pertinent ring. For any terphenyl (Kekulé form), the total



term is 24 (the sum of 8 + 8 + 8), while for any cyclohexylbiphenyl, it is 17 (the sum of 1 + 8 + 8).

The term $\sum (-)^2_a (-)^2_b$ has values of 0, 0, 4, 36, 2, 5, and 52 for the hydrogen-suppressed² molecules C—C,

C—C—C, C=C—C—C, C=C—C≡C, C—C, C=C, C—C=C—C≡C—C, respectively, since contributions are counted once only from the squared products of non-adjacent bond orders. In other words, for the molecule

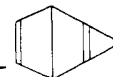


where the letters *j*, *k*, *l*, *m*, and *n* denote the bond orders for the connectivities pictured, $\sum (-)^2_a (-)^2_b$ equals $j^2l^2 + j^2m^2 + k^2m^2 + l^2n^2$.

A more detailed example of ACMCP computation by geometric means is shown in Table II for a complex hypothetical homoatomic ring molecule containing seven carbon atoms in a tricyclic system. The terms are shown combined to give the hydrogen-suppressed ACMCP.² If the necessary number of hydrogen atoms are included in the first term, the resulting modified hydrogen-suppressed ACMCP begins with C⁷H⁶, the simple molecular formula for the molecule. Of course, the same ACMCP can be computed from the appropriate ACM by the usual manual or machine methods.³

The coefficients in Table I can be extended to terms of degree lower than Xⁿ⁻⁷ by rules based on extrapolative

Table II
Values of Coefficient Terms for the Molecule



$$\sum (-)^2 = 1^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 1^2 = 15$$

$$\sum \triangle = 2 \cdot 1 \cdot 1 = 2$$

$$\sum \square = 2 \cdot 1 \cdot 1 \cdot 1 + 2 \cdot 1 \cdot 1 \cdot 1 = 4$$

$$\sum (-)^2_a (-)^2_b = 2^2 \cdot 1^2 + 2^2 \cdot 1^2 + 2^2 \cdot 1^2 + 2^2 \cdot 2^2 + 2^2 \cdot 1^2 + 2^2 \cdot 1^2 + 1^2 \cdot 1^2 + 1^2 \cdot 1^2 + 1^2 \cdot 2^2 + 1^2 \cdot 1^2 + 1^2 \cdot 1^2 + 1^2 \cdot 2^2 + 1^2 \cdot 1^2 + 1^2 \cdot 1^2 + 1^2 \cdot 1^2 + 1^2 \cdot 1^2 = 60$$

$$\sum \odot = 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 = 1$$

$$\sum (-)^2_a \triangle_b = 1^2(2 \cdot 1 \cdot 1) + 1^2(2 \cdot 1 \cdot 1) + 1^2(2 \cdot 1 \cdot 1) + 2^2(2 \cdot 1 \cdot 1) = 14$$

$$\sum \odot = 2 \cdot 1 \cdot 1 \cdot 2 \cdot 1 \cdot 1 = 4$$

$$\sum (-)^2_a \square_b = 2^2(1 \cdot 1 \cdot 2 \cdot 1) + 2^2(1 \cdot 1 \cdot 2 \cdot 1) + 1^2(1 \cdot 1 \cdot 2 \cdot 1) + 1^2(1 \cdot 1 \cdot 2 \cdot 1) = 20$$

$$\sum \triangle_a \triangle_b = 0$$

$$\sum (-)^2_a (-)^2_b (-)^2_c = 1^2 \cdot 1^2 \cdot 2^2 + 1^2 \cdot 1^2 \cdot 1^2 + 1^2 \cdot 1^2 \cdot 1^2 + 1^2 \cdot 1^2 \cdot 1^2 + 2^2 \cdot 1^2 \cdot 2^2 + 2^2 \cdot 1^2 \cdot 1^2 + 2^2 \cdot 1^2 \cdot 1^2 + 2^2 \cdot 1^2 \cdot 1^2 + 2^2 \cdot 1^2 \cdot 1^2 + 2^2 \cdot 1^2 \cdot 1^2 = 44$$

$$\sum \odot = 1 \cdot 2 \cdot 1 \cdot 1 \cdot 1 \cdot 1 \cdot 1 = 2$$

$$\sum (-)^2_a \odot_b = 2^2(1 \cdot 1 \cdot 1 \cdot 1 \cdot 1) = 4$$

$$\sum \triangle_a \odot_b = (2 \cdot 1 \cdot 1)(2 \cdot 1 \cdot 1 \cdot 1) = 4$$

$$\sum (-)^2_a (-)^2_b \triangle_c = 2^2 \cdot 1^2(2 \cdot 1 \cdot 1) + 1^2 \cdot 1^2(2 \cdot 1 \cdot 1) = 10$$

$$\therefore \text{TOTAL HYDROGEN-SUPPRESSED ACMCP} = C^7 - 15C^5 + 4C^4 + 52C^3 - 26C^2 - 12C$$

induction from the values already given. For example, the coefficient for the X^{n-k} component of an ACMCP is a linear combination of all possible different bond order (connectivity) products of degree k , with the provisos that all isolated bond orders must be squared, that all rings must be closed, and that all geometric factors in a given term must be nonadjacent. [(A "geometric factor" is defined as being a ring or an isolated bond (a two-membered ring).] Each term is multiplied by a factor of two for each ring in its makeup because of two different possible directions of traversal, and the sign of each term of degree X^{n-k} is determined by $(-1)^{k-j}$, where j is the number of geometric factors in the term. For example, in the coefficient to X^{n-12} , there is a term of the form

$$-8 \sum \text{[Diagram: a hexagon with a triangle inside, labeled a, b, c]}_a \text{[Diagram: a triangle, labeled b]}_b \text{[Diagram: a triangle, labeled c]}_c$$

and another one,

$$-4 \sum \text{[Diagram: a hexagon with a square inside, labeled a, b]}_a \text{[Diagram: a square, labeled b]}_b (-1)^2_c,$$

and yet another,

$$+8 \sum \text{[Diagram: a square with a triangle inside, labeled a, b, c]}_a \text{[Diagram: a triangle, labeled b]}_b \text{[Diagram: a triangle, labeled c]}_c (-1)^2_d,$$

to name just three.

For a molecule of m atoms, k is always less than, or equal to, m . It can never be greater than m .

It is instructive to note that the "total degree"⁷ of every term in the ACMCP for a molecule composed of n atoms is n . In other words, the product of connectivity factors and atomic factors (the lettered symbols) for any given term of an ACMCP for an n -particle molecule is always of the n th degree in its factors. With this significant information, we are now in a position to understand the heteroatomic ACMCP.

The Heteroatomic ACMCP.—The ACMCP for the heteroatomic case differs from the homoatomic case discussed above only in diversity of the lettered factors (atomic symbols) for each term. The numerical coefficients still have the same physical significance in being combinatorial products of connectivities in bonds and around ring configurations in the molecule under consideration.

To understand the nature of the atom symbol combinations it proved valuable to evaluate the ACMCP for a completely generalized (but symmetric, to preserve molecular significance) ACM of order seven (*i.e.*, a seven-atom molecule) wherein *all* atoms differ from each other. The ACM can be written as

A	a	b	c	d	e	f
a	B	g	h	i	j	k
b	g	C	l	m	n	o
c	h	l	D	p	q	r
d	i	m	p	E	s	t
e	j	n	q	s	F	u
f	k	o	r	t	u	G

This matrix is equivalent to the generalized seven-centered heteroatomic interconnectivity model depicted in

(7) By "total degree" of a term is meant the sum of exponents of the lettered factors (the usual term degree) plus the number of bond connectivity factors (the bond degree) in the numerical coefficient. For example, for the term

the usual term degree is $(n-12)$, the bond degree is $4+3+3+2=12$, and the "total degree" is $n-12+12=n$.

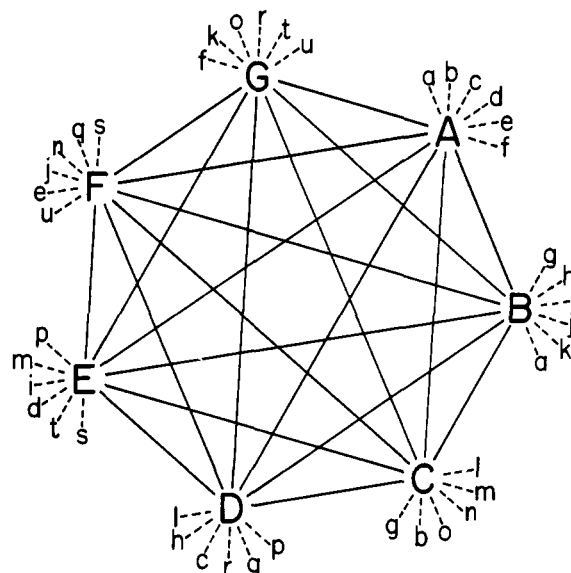


Fig. 2.—Schematic pictograph of generalized seven-centered heteroatomic interconnectivity model.

Fig. 2. In this figure, the atoms, A, B, C, D, E, F, and G are shown as the foci, and the connectivities between respective pairs are identified by lower case letters at both ends of the lines (extended). For example, the connectivity between atoms B and D is h , in agreement with the ACM.

Expansion of the subject ACM into its ACMCP yields many more terms than does the homoatomic case since the atoms now are distinguishable. For each atom term of degree $n-k$, the coefficients, which are product combinations of connectivities, are still of the form given in Table II. However, the atom factor X^{n-k} is a different product combination of the atoms for each connectivity combination. This can best be understood by examining specific terms of the generalized heteroatom ACMCP shown in Table III.

In most of Table III, only a few representative terms are given, but if one traces out the connectivities for each term with Fig. 2 as a guide, the same geometric groups are obtained as in the analogous case of the homoatomic system (Table I). However, *each geometric group is multi-*

Table III

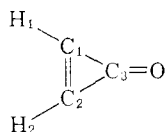
Term degree (in atoms)	Complete term
7	ABCDEFG
6	0
5	$-(a^2CDEFG + b^2BDEFG + c^2BCEFG + \dots + u^2ABCDE)$
4	$+2(abgDEFG + achCEFG + cdpBCFG + \dots + efuBCDE)$
3	$-[2(acglEFG + adgmDFG + \dots + efqrBCE) - (a^2s^2CDG + b^2i^2DFG + \dots + d^2u^2BCD)]$
2	$+ [2(adglpFG + acirtCF + \dots + gjoqrAE) - 2(a^2qruCE + a^2lorEF + a^2mnsDG + \dots + u^2adiCD)]$
1	$-[2(aeglpG + hinorsA) - 2(a^2noqrE + \dots + u^2bdlpB) - 4(abg \cdot prtF + \dots + adi \cdot lorF) + (a^2l^2s^2G + \dots + a^2n^2r^2E)]$
0	$+ [2(afglpsu + dflnpu + \dots + cdhjinot) - 2(a^2lopsu + a^2noprs + \dots + u^2bchim) - 4(adi \cdot loqu + adi \cdot lnru + \dots + ceq \cdot giot) + 2(a^2o^2pqs + \dots + a^2n^2prt)]$

plied by all of the other atoms not involved in its connectivity factors. For example, in Fig. 2, if the only geometric group in a given term is the triangular pattern *lor* (which utilizes atoms C, D, and G) then this is multiplied by the product of the other atoms, ABEF.

Based on this general principle, the first term of the ACMCP, since it utilizes no connectivities, must include the product of all the atoms in a molecule, hence giving the molecular formula. The second term of the ACMCP would require a connectivity involving only one atom, but since this is physically impossible in the present case, the coefficient has a value of zero. The third term of the ACMCP involves the square of a connectivity which ties up two atoms. This leaves $n - 2$ atoms as factors to account for the lettered degree of this term. This same mode of construction applies down through all the other terms of the ACMCP in appropriate fashion; even to the constant term in which, because the molecular connectivities utilized in the geometric groups involve all atoms in the molecule, no lettered atomic factors are available and the atomic degree is zero.

This result explains why so much simplification is introduced when hydrogen atoms are suppressed or omitted. The more bonds or connectivities that an atom has in a molecule, the more likely it is that one or more of these bonds will be involved in one of the coefficients of the ACMCP and its atom symbol will not then be used. However, hydrogen, by virtue of its monovalency and its almost universal omnipresence in organic molecules is likely to "clutter" the ACMCP.

This situation as well as the applicability of the rules for the heteroatomic case are exemplified by the following example, utilizing the hypothetical molecule, cyclopropenone.



In the structure, the atoms which appear more than once are separately numbered to assist in tracing the analysis of the ACMCP.

The first term of the ACMCP for this molecule is $C_1C_2C_3H_1H_2O$. This is of atomic degree 6. The term in atomic degree 5 is zero.

The next term, of fourth degree in the atoms, is the following:

$$-(1 \cdot 1)C_2C_3H_2O + (2 \cdot 2)C_3H_1H_2O + (1 \cdot 1)C_1C_3H_1O + \\ (1 \cdot 1)C_3H_1H_2O + (1 \cdot 1)C_1H_1H_2O + (2 \cdot 2)C_1C_2H_1H_2]$$

The term in degree 3, ($n - 3$ equals $6 - 3$) is $+2[(2 \cdot 1 \cdot 1) \cdot H_1H_2O]$. This is followed by the next term $[(1^2 \cdot 1^2)H_2O + (1^2 \cdot 2^2)C_2H_2 + (1^2 \cdot 1^2)C_3O + (1^2 \cdot 1^2)H_1O + (1^2 \cdot 2^2)C_1H_1 + (2^2 \cdot 2^2)H_1H_2]$. The term of atomic degree 1 has a zero coefficient since there is neither a five-membered ring nor a three-membered ring with a nonadjacent bond. The constant term receives contribution only from the squared product of three mutually nonadjacent bonds (which involve all the atoms) and thus equals $1^2 \cdot 1^2 \cdot 2^2 = 4$.

If the subscripts are ignored

$$\text{ACMCP cyclopropenone} = C^3H^2O - (2C^2HO + 6CH^2O + \\ 4C^3H^4) + 4H^2O + 2HO + 8CH + CO + 16H^2 + 4$$

This may be compared with the modified hydrogen-suppressed ACMCP,⁸ $C^3H^2O - (6CO + 4C^2) + 4O + 16$; $C(1,2,0,0)$. The latter has only one-half the number of terms of the former. It is interesting to note that the only coefficients of the former that appear in the latter are those which multiplied atomic terms containing H^2 as a factor. In general, for a molecule containing h hydrogen atoms, the hydrogen-suppressed ACMCP retains only those terms of the total ACMCP which contain H^h in them (although these powers of H are subsequently dropped). This observation is understandable since all terms of lower powers in H than h , must carry a factor within their coefficients involving at least one connectivity to H . Since such connectivities are all considered to be zero for the hydrogen-suppressed case, these coefficients must vanish.

Attributes of the ACMCP.—From the expansions of the general ACM's discussed above it is evident that the resulting ACMCP's, whether expressed in algebraic or geometric form, are independent of the order of placement into the diagonals of the ACM. This is true since each term of the ACMCP consists only of atoms and connectivities as factors. This implies that the atoms are identified only by their connectivities to their nearest neighbors, and then these to their neighbors, and so on. In other words, the ACMCP has its own internal self-consistent coordinate system, and this is the same as that of the pictograph and that of the ACM. Although this internal consistency is less obvious in the ACM because of the seemingly different permutations which can be constructed, it must still be present because of the retention of the single-valued mapping from the pictograph into the ACMCP through the ACM as an intermediate.

The ACMCP is indeed a philosophically satisfying construct. It has gratifying symmetry in that each of the terms is of constant total degree in connectivities and atoms so that the polynomial is a type of microcosm. Each term represents a "state" of the microcosm wherein some atoms participate, through their connectivities, to the value of a coefficient, whereas the others, the non-participants for the nonce, are lumped together in the atoms factor. The aggregate sum of such "states" is the molecule.

Other Applications.—The ability of the physico-geometric interpretation of the terms in the ACMCP to permit direct calculation of the polynomial from the pictograph without going through the matrix suggests that the ACM may be dispensed with. This could well be so for the simpler molecules but for computer calculations in the case of large and complex molecules, the ACM may still have utility. However, the above enunciated rules do permit ready evaluation of the first few terms of the ACMCP, *i.e.*, a truncated ACMCP,² which can form the basis for a manual filing system.

The physico-geometric rules also allow computation of the secular equation in simple LCAO molecular orbital calculations on p-atomic orbital systems without setting up the usual matrix. One simply draws the molecule as a series of adjacent p-atomic orbitals and computes the ACMCP with the recognition that the unit of connectivity is the resonance exchange integral (β). In the resulting ACMCP, one replaces each atom symbol by its appropri-

(8) The valency count, here added in the form $C(1,2,0,0)$, is required to prevent polynomial degeneracy as explained elsewhere.²

ate $(-\lambda + c)$ where c is some numerical correction for electronegativity and/or σ -bond energy. The resulting polynomial is a function of λ , the usual energy eigenvalue, and when equated to zero is identical with the traditionally obtained secular equation in appropriate energy units.

SUMMARY

The physico-geometric representation of the ACMCP is described by the expression

$$ACMCP = \sum_{k=0}^N \left\{ (-1)^{\sum_{i=1}^{N-k} n_i} \cdot (2)^{\sum_{i=1}^{N-k} n_i} \cdot \left[\prod_{j=2}^{N-k} \prod_{m=0}^{n_j} Q_m^j \right] S_k \right\}$$

where \sum_{*} = summation over all different combinations of non-adjacent rings where

$$\sum_{i=1}^{N-k} i n_i = N - k$$

"combination" = collection of nonadjacent rings whose connectivities enter into evaluation of a particular coefficient in ACMCP

i = integer, $2 \leq i \leq N - k$

j = integer, $2 \leq j \leq N - k$

k = integer, $0 \leq k \leq N$; identifies degree of ACMCP term

m = integer, $0 \leq m \leq n_j$

n_i = integer, $0 \leq n_i \leq [(N - k)/i]$, where upper limit is the largest integer value of $(N - k)/i$; represents number of nonadjacent (nontouching) i -membered rings in a combination

Q_m^j = product of bond connectivities around (defining) the m th nonadjacent j -membered ring

S_k = product of atomic symbols, k in number, which are *not* included in the rings making up a given combination

S_0 = 1, for any j ; a definition

Q_0^j = 1; a definition

Acknowledgment.—Acknowledgment is gratefully made to referee Dr. Sylvan Eisman, Pitman-Dunn Institute for Research, U. S. Army, Frankford Arsenal, Pa., who called the author's attention to a paper by F. Harary, [*J. Math. Phys.*, 38, 104 (July, 1959)] which presents a graph-theoretic study anticipating some of the relations independently found here. However, it is couched in mathematical terminology not normally understood by the average chemist and does not recognize the significance of explicit atom (vertex) labeling so important for molecular pictographs, particularly the heteroatomic ones.

Chemical Structure Fragmentation for Use in a Coordinate Index Retrieval System*

By RAY W. IHNDRIIS

Cancer Chemotherapy National Service Center, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland

Received January 28, 1964

This paper describes the changes that have been made in the Cancer Chemotherapy National Service Center (CCNSC) Chemical Information Retrieval System¹ established for analog searching of chemical structures.

The National Bureau of Standards Peek-a-Boo System² in use at the CCNSC consists of a precision-built machine which uses a microswitch to actuate the solenoid-operated punch when the plastic card is placed into punching position. The cards used for permanently recording "data" are 5 × 8 in., 8–10 mil vinyl plastic sheets. Each card is punched separately.

The Reader consists of a back-lighted opaque frame onto which the plastic cards are placed and over-layed with a grid so that the coincident holes allow light to shine through, and file numbers can be "read" from the coordinate position on the grid. The thousands and hundreds are read along the horizontal axis, and the tens

and units are read on the vertical axis. The grid provides for 18,000 recording positions.

A VISIrecord file cabinet houses the plastic cards in a manner permitting easy access. The cards are shingled so that three-fourths of an inch along the top edge exposes the chemical fragment or other descriptor. The cards are notched on the right-hand edge and filed vertically so that they fit over rods in the bottom of the file cabinet. They are printed to define the space for the descriptors and also have blocks along the left-hand edge that signal the position when a card is removed. Thus it is easy to find a card for entering new data or retrieving information.

One or several descriptors are used for selection by overlaying the plastic cards, thus eliminating unwanted information. We can vary the search at any time if the desired information is not found, or conversely, if too much chaff appears.

The improvements in our system of coding chemical structures affect the ring systems, the fragment chains, and the supplier code.

* Presented before the Division of Chemical Literature, 145th National Meeting of the American Chemical Society, New York, N. Y., Sept. 11, 1963.