CHEMICAL GRAPHS—V

ENUMERATION AND PROPOSED NOMENCLATURE OF BENZENOID *CATA*-CONDENSED POLYCYCLIC AROMATIC HYDROCARBONS[1]

A. T. BALABAN and F. HARARY[2]

Institute of Atomic Physics, Bucharest, Roumania; University of Michigan, Ann Arbor, U.S.A.; and University College, London

Abstract—A modified definition is proposed for *cata*-condensed aromatic hydrocarbons. A formula is devised for the number of isomers of non-branched *cata*-condensed arenes in function of the number of hexagons. Proposals are made for a graph-systematic nomenclature of such arenes, based on the characteristic tree of these systems.

THE nomenclature of benzenoid condensed polycyclic aromatic hydrocarbon is extremely complex. In Clar's two-volume monograph[3] the summary extends over 13 + 48 pages, since the German nomenclature system adopted therein is different from the IUPAC,[4] Chemical Abstracts[5] or Ring Index[6] systems, and it was necessary to depict the formulae of all compounds discussed in order to indicate rapidly and unambiguously which compounds were discussed. This fact illustrates the basic faults of all nomenclature systems adopted so far in the class of benzenoid polycyclic hydrocarbons. Trivial names (coronene, pyrene, ovalene, violanthrene, zethrene, chrysene, picene, circumanthracene, perylene, terrylene, anthanthrene, bisanthene, peropyrene, (iso)violanthrene, pyranthrene, etc.) along with semisystematic names (acenes, phenes, helicenes, rylenes) and with different alternative systems for indicating by numbers[3] or letters[4-6] the homology (benzo, naphtho, etc.) make the nomenclature extremely confusing.

No attempt has apparently been made to devise a formula for counting the various isomeric polycyclic aromatic hydrocarbons. In the present paper we indicate a counting series for a part of the polycyclic hydrocarbons, proposing at the same time a simple nomenclature system. This may be useful for computer handling of literature data (cf.[7]), taking into account that many systems falling under the proposed nomenclature have important applications in cancer research,[8] in the dyestuff industry,[9] and in theoretical, physical and organic chemistry.[3, 10]

All benzenoid polycyclic aromatic hydrocarbons (arenes) are superimposable with the graphite lattice. They are classified into several groups as shown in Fig. 1.* Condensed (fused†) polycyclic arenes have two adjacent carbon atoms common to two rings, like naphthalene I.

---

* In formulas I–XV, double bonds or inscribed circles in aromatic rings have been omitted for clarity.
† In the IUPAC 1957 Rules,[4] *cata*-condensed systems are named *ortho*-fused, and *peri*-condensed systems are named *ortho*- and *peri*-fused.
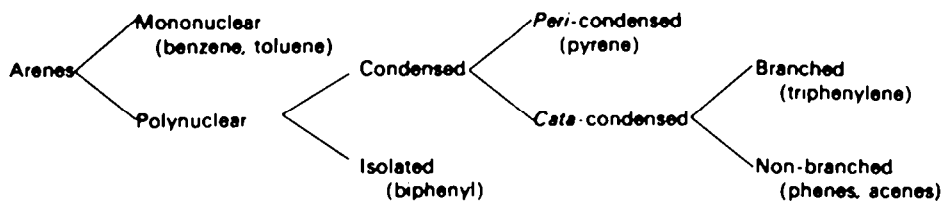
FIG. 1 Classification of arenes (examples are given in brackets).

The usual criterion for classifying a condensed polycyclic arene as *cata-* or *peri-*condensed is the absence or presence of a C atom common to three rings. We propose a modified, graph-theoretical definition of *cata-* or *peri-*condensed systems. If the centers of the hexagons are joined by a line when two hexagons are condensed, one obtains a connected graph which will be named "characteristic graph" or "skeleton". By the new, proposed, definition if this graph is a tree, the arene is *cata-*condensed; if this graph is not a tree (if it contains rings), the arene is *peri-*condensed.
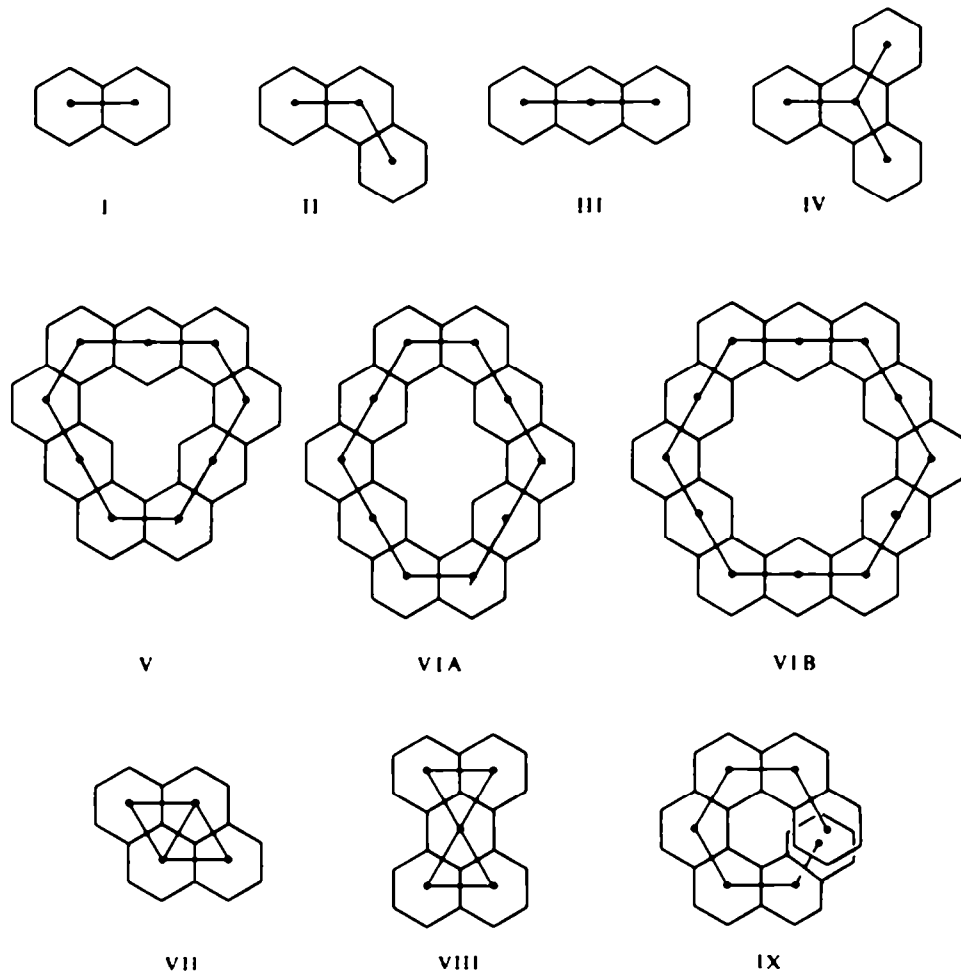
Systems I–IV are classified as *cata*-condensed both by the usual and by the new definition; likewise, pyrene VII and perylene VIII are *peri*-condensed systems in both definitions. The two definitions are, however, at variance in the case of systems such as V, VIA[11a] and VIB,[11b] which are *cata*-condensed in the usual definition, but *peri*-condensed in the new one. We favour the new definition because if this is accepted, all *cata*-condensed arenes with the same number n of hexagons are isomeric, i.e. have the same molecular formula $C_{4n+2}H_{2n+4}$. This is not true for *peri*-condensed systems, both in the usual and in the new definitions.

Compound V, $C_{36}H_{18}$ is not an isomer of nonacene or of other *cata*-condensed hydrocarbons with nine hexagons $C_{38}H_{22}$, neither are VIA, $C_{40}H_{16}$, and VIB, $C_{48}H_{24}$, isomers of decacene, $C_{42}H_{24}$ and dodecacene, $C_{50}H_{28}$, respectively; similarly the *peri*-condensed system pyrene VII, $C_{16}H_{10}$ is not an isomer of tetracene $C_{18}H_{12}$, neither is perylene VII, $C_{20}H_{12}$, an isomer of pentacene $C_{22}H_{14}$. Accordingly, it is reasonable to classify V, VIA, VIB and other coronaphenes as *peri*-condensed systems.*

Compounds like hexahelicene or the recently prepared[12] heptahelicene IX which are non-planar and cannot be drawn on paper (therefore they are not strictly superposable with the graphite lattice) can be considered as *cata*-condensed systems.

By adopting henceforth the new definition of *cata*- and *peri*-condensed systems, we may hope to be able to devise a formula giving the number C_n of the isomers of a given *cata*-condensed system with n hexagons. The solution of this problem will be presented in this paper together with a new nomenclature system of *cata*-condensed arenes.

This also provides a partial answer to the hexagonal cell-growth problem, which was repeatedly mentioned by one of us[13–16] as one of the unsolved problems in the theory of graphs. This problem, which can be formulated: "how many hexagonal-cell aggregates or 'animals' are there for a given number n of hexagons", is equivalent to counting all condensed (*cata* and *peri*) arenes with n hexagons.

By empirical approaches the condensed polycyclic arenes shown in Fig. 2 and presented in Table 1 have been found.

The *peri*-condensed systems present many complicating features: when their characteristic graph has an odd number of triangles, they are radicals (r in Table 1) like perinaphthyl X, or naphtanthryl XI; with an even number of triangles, they can be diradicals (d) like coeranthrene XII, or like XIII and XIV, or can be "normal" (p) like pyrene VII and perylene VIII. In the former case, it is impossible to accommodate one double bond at every carbon atom in the sense of Kekulé structures. If all atoms are distributed into two alternate sets of starred and non-starred atoms,[17] the sets

---

* Alternatively, one could conserve the previous meaning of *peri*-condensed systems by ascribing this name to compounds whose skeleton contains three-membered rings (triangles), and by coining for compounds such as V or VI, whose skeleton has a larger ring, a new name, e.g. *corona*-condensed systems; besides other interesting features such as ring current or clathrate formation, these latter systems could lead to Möbius-strip molecules, if two adjacent bonds change places.

X   XI   XII

XIII   XIV   XV
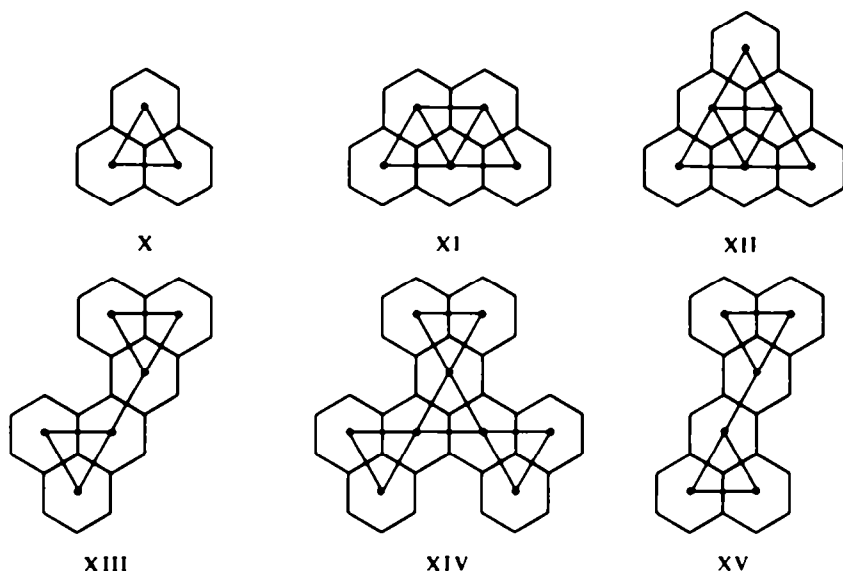
have equal numbers of atoms in normal compounds and unequal in diradicals.[*] Structural features (z) like those present in zethrene XV determine the localization of the double bonds shown in the formula. Since the Hückel rule of aromaticity breaks down in *peri*-condensed aromatic systems it is mostly in *cata*-condensed compounds that Kekulé structures are important, as evidenced by Clar[18] or by correlations[19] between the maximum number of rings in which three alternate double bonds can be written and the delocalization energy.

Starting from the observation[20] that in the graphite lattice all lines have one of three directions which can be noted by digits ± 1, ± 2, ± 3 similarly to the four-direction problem in the diamond lattice,[21] one can describe the outline of a condensed arene by a number consisting of these digits, e.g. benzene by $1\bar{3}2\bar{1}3\bar{2}$ and
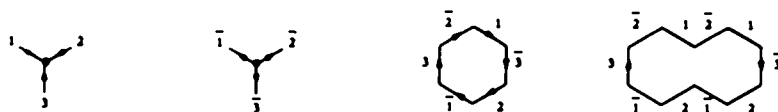

Fig. 3

naphthalene by $1\bar{3}2\bar{1}2\bar{1}3\bar{2}1\bar{2}$ (Fig. 3). This approach seems suitable for computer use. These numbers must have several restrictions: every positive digit must have in the number its negative counterpart; two adjacent digits must have opposite signs; only sequences, $1\bar{2}$, $1\bar{3}$, $2\bar{1}$, $2\bar{3}$, $3\bar{1}$, $3\bar{2}$, and the same with reversed signs, are allowed.

A simpler approach is to consider the characteristic graph associated to the arene. Owing to the complications introduced by *peri*-condensed systems, we shall discuss in the following only *cata*-condensed compounds. Their graph is a tree, which may be branched or non-branched; in the latter case the tree has only two endpoints.

---

[*] Triangular arrangements of hexagons like X, XII, etc., lead to monoradicals, diradicals, ..., q-radicals. Such a q-radical is a triangular array of $\frac{1}{2}(q + 1)(q + 2)$ hexagons (one side of the equilateral triangle consists of $q + 1$ hexagons), containing $\frac{1}{2}(q + 6)(q + 1)$ starred and $\frac{1}{2}(q + 3)(q + 2)$ non-starred atoms.
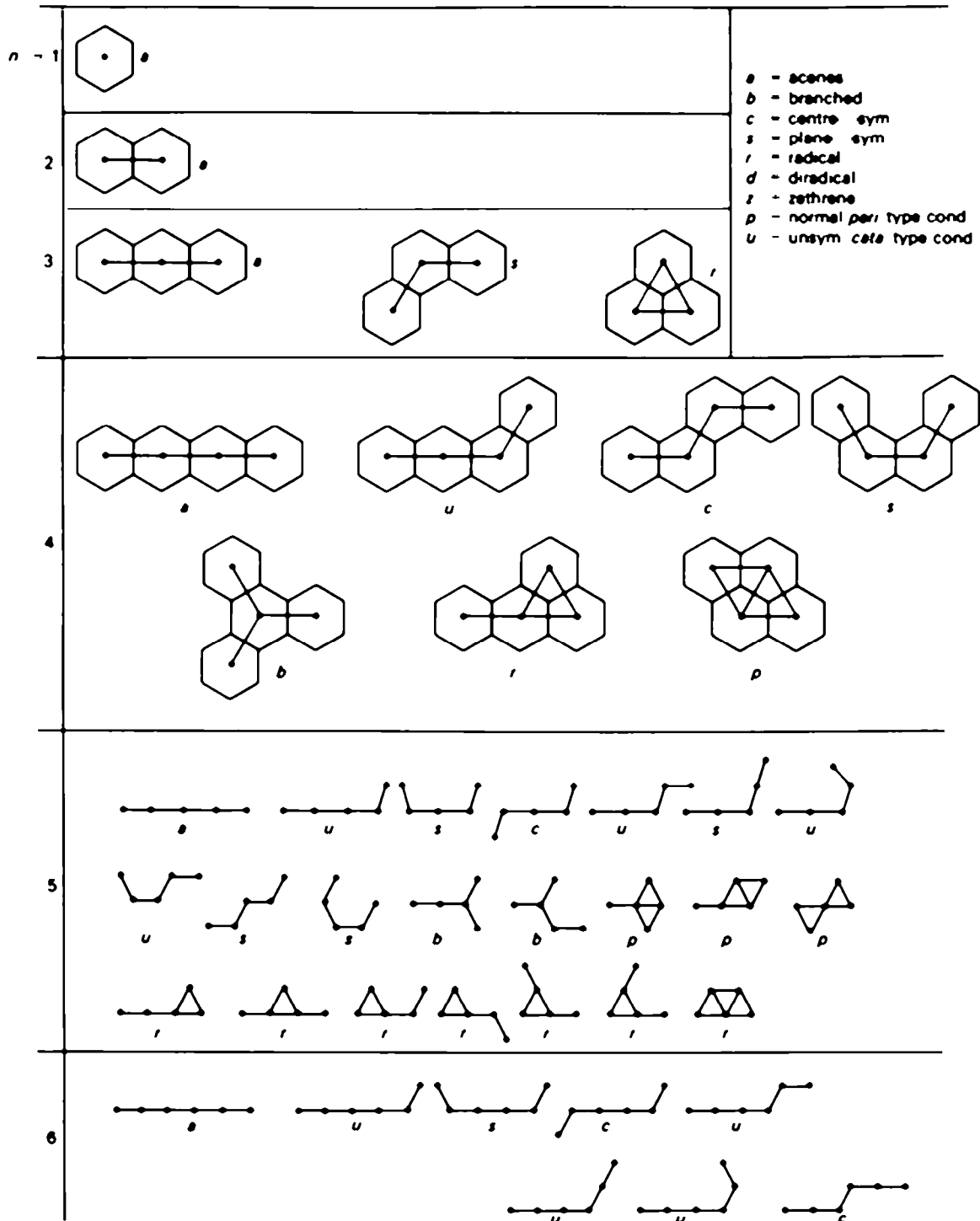
Fig. 2 Condensed polycyclic arenes

a — acenes
b — branched
c — centre sym
s — plane sym
r — radical
d — diradical
z — zethrene
p — normal peri type cond
u — unsym cata type cond

FIG. 2—*continued*

and other *peri*-condensed hydrocarbons
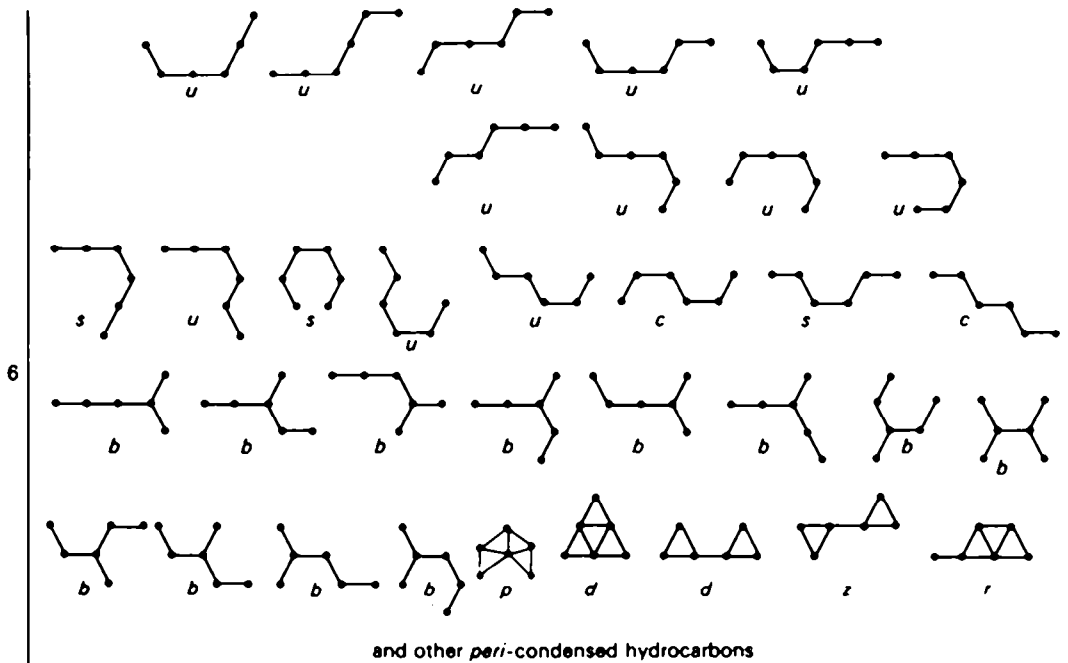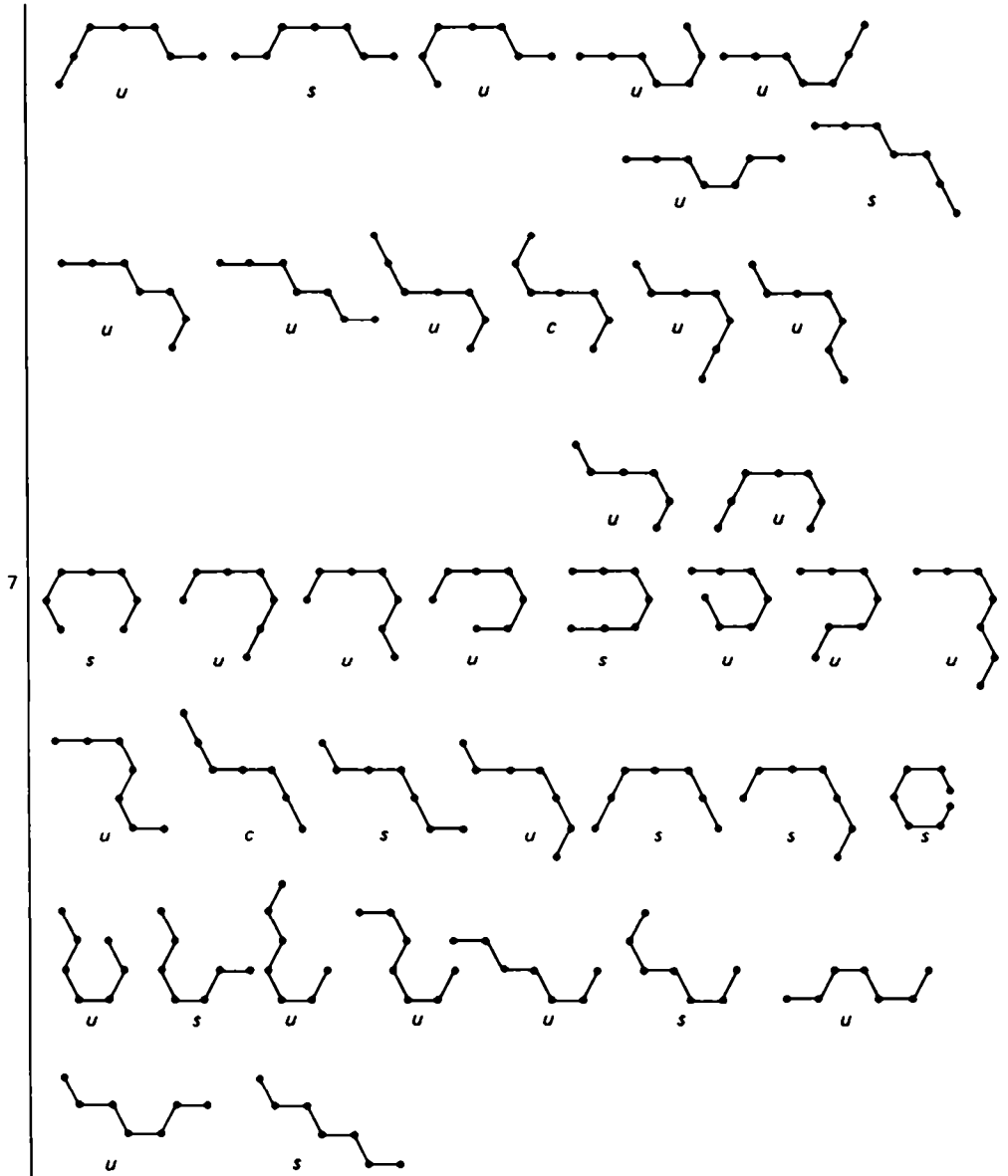
7

and also branched *cata*-condensed and *peri*-condensed hydrocarbons

TABLE 1. EMPIRICAL NUMBERS OF ISOMERS OF CONDENSED POLYCYCLIC ARENES

No. of hexagons	Cata-condensed						Peri-condensed					Grand total
	Non-branched					Branched					Total	
		1	2	4	Total							
n	a	s	c	u	C_n	b	p	r	d	z	P_n	c + b + P
1	1	—	—	—	1		—	—	—	—		1
2	1	—	—	—	1	—						1
3	1	1			2	—	1	—	—	—	1	3
4	1	1	1	1	4	1	1	1			2	7
5	1	4	1	4	10	2	3	7	—		10	22
6	1	4	4	4	25	12	14	28	2	1	45	82[c]
7	1	13	4	52	70	[a]	[a]	[a]	[a]	[a]	[a]	333[c,d]
8	1	13	13	169[b]	196[b]	[a]	[a]	[a]	[a]	[a]	[a]	1448[d]

[a] Not determined.
[b] Calculated after Eqs (7b) and (8b).
[c] Results of computer calculations by L. S. Kassel (pers. comm.)
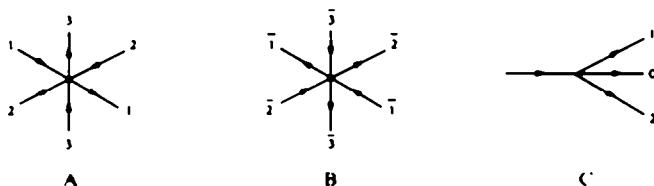[d] Personal communication from Dr. Martin Gardner.

FIG. 4.

The present discussion is limited to non-branched, cata-condensed systems. Their associated skeleton (characteristic) trees differ from ordinary trees in that a kinked and a straight line are not equivalent (cf. II and III). The directions of the lines in the graph being the same as before (Fig. 4A and B), the tree may be described by a number consisting of the digits ±1, ±2, ±3 with the restriction that only sequences 11, 1$\bar{2}$, 1$\bar{3}$, 2$\bar{1}$, 22, 2$\bar{3}$, 3$\bar{1}$, 3$\bar{2}$, 33, and the same with reversed signs, are allowed. An alternative simpler notation of the tree makes use of the observation that two lines incident to the same point can make angles of 120°, 180° or 240°. Two adjacent lines with an angle of 180° will be denoted by 0, while two adjacent lines with angles 120° or 240° will be denoted by 1 or 2 (Fig. 4C). The characteristic tree of a non-branched cata-condensed arene with n hexagons (which is a directed path from one endpoint to the other) will therefore be denoted by a number of $n - 2$ digits formed from digits 0, 1 and 2. Examples of such numbers are: 1 or 2 (for II), 0 (for III), 11111 or 22222 (for IX).

One isomer may be indicated by one, two or four numbers, e.g. the above examples for II, III and IX, and 01, 02, 10 or 20 for tetraphene. Acenes are the only compounds whose tree is a straight line, indicated by one number 00 . . . 00; in Table 1 and Fig. 2 they are indicated by a. Symmetrical trees, either relatively to an axis (they are noted by s) or to a center (noted by c), are described by two numbers which differ by

interchanging 1 with 2; their endpoints are equivalent, i.e. the same number is obtained if any of them is the starting point. Unsymmetrical trees (denoted by u) are described by four numbers differing by the starting point or by interchanging digits 1 and 2. The heading of Table 1 indicates for non-branched *cata*-condensed arenes: 1 number for a, 2 for s or c, and 4 for u. Since there are 3^{n-2} numbers formed from $n - 2$ digits 0, 1 or 2, which must indicate all C_n isomers of a non-branched *cata*-condensed arene composed of n hexagons, one can write equation (2). Eqs 1 and 3 are straightfoward.

$$C_n = a_n + s_n + c_n + u_n \tag{1}$$

$$3^{n-2} = a_n + 2(s_n + c_n) + 4u_n \tag{2}$$

$$a_n = 1 \tag{3}$$

$$s_{2m-1} = \tfrac{1}{2}(3^{m-2} - 1) \quad \text{for } n \text{ odd}, n = 2m - 1 \tag{4a}$$

$$s_{2m} = \tfrac{1}{2}(3^{m-1} - 1) \quad \text{for } n \text{ even}, n = 2m \tag{4b}$$

$$c_{2m-1} = \tfrac{1}{2}(3^{m-1} - 1) \quad \text{for } n \text{ odd}, n = 2m - 1 \tag{5a}$$

$$c_{2m} = \tfrac{1}{2}(3^{m-1} - 1) \quad \text{for } n \text{ even}, n = 2m \tag{5b}$$

Equations (4) and (5) can be demonstrated by induction. Symmetrical trees (presented in Fig. 5), s_n or c_n are obtained from every symmetrical tree s_{n-2} and c_{n-2}, respectively, by adding two lines at the endpoints at angles 120°, 180° or 240°; one more tree is derived from the acene a_{n-2}; therefore we have the recurrence formulas (6), which with Fig. 5 gives Eqs (4a) (5b).

$$s_n = 3s_{n-2} + 1; \qquad c_n = 3c_{n-2} + 1 \tag{6}$$

From Eqs (1)–(5b) we obtain:

$$u_{2m-1} = \tfrac{1}{4}(3^{2m-3} - 3^{m-1} - 3^{m-2} + 1) \quad \text{for } n \text{ odd}, n = 2m - 1 \tag{7a}$$

$$u_{2m} = \tfrac{1}{4}(3^{m-1} - 1)^2 \quad \text{for } n \text{ even}, n = 2m \tag{7b}$$

$$C_{2m-1} = \tfrac{1}{4}(3^{2m-3} + 3^{m-1} + 3^{m-2} + 1) \quad \text{for } n \text{ odd}, n = 2m - 1 \tag{8a}$$

$$C_{2m} = \tfrac{1}{4}(3^{m-1} + 1)^2 \quad \text{for } n \text{ even}, n = 2m \tag{8b}$$

The number C_n of isomers calculated by means of formulas (8a) and (8b) gives results consistent with Table 1.

The nomenclature of non-branched *cata*-condensed arenes can be simply derived from the characteristic (skeleton) graph: the smallest number denoting the skeleton, in brackets (this number has $n - 2$ digits), is followed by the Greek numeral indicating the number n of hexagons with the suffix catafusene. Some graph-systematic names are indicated in Table 2. They are easily converted into a pictorial representation by drawing the tree (1 and 2 mean left or right, 0 means straight on) and then completing the hexagonal frame. These names are of practical value for $n \geqslant 4$.

In the future these ideas will be applied to branched *cata*-condensed arenes, and possibly also to *peri*-condensed systems.
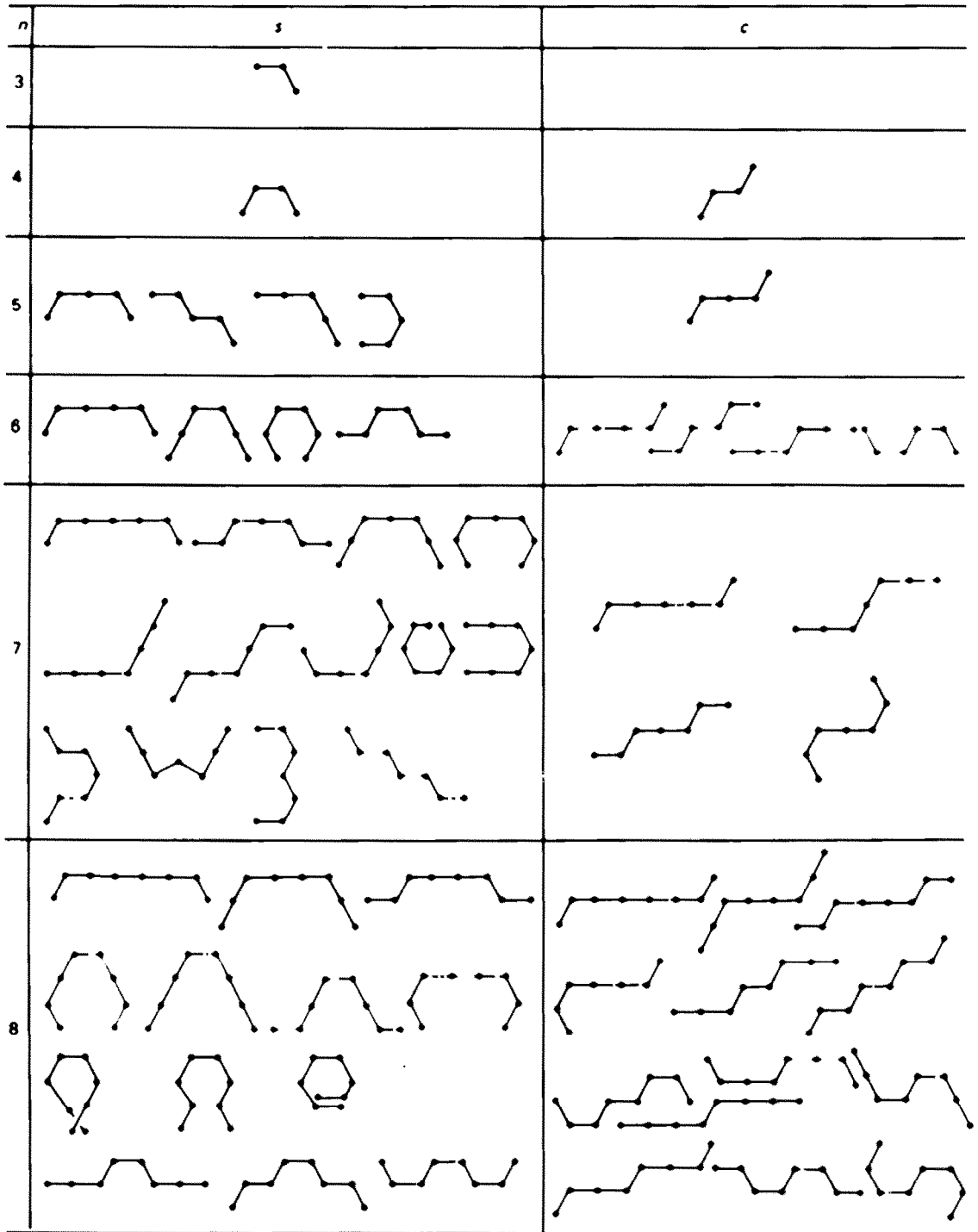
n	s	c

FIG. 5 Symmetrical trees of types s and c

TABLE 2. CORRESPONDENCE BETWEEN TRIVIAL AND GRAPH-SYSTEMATIC
NAMES OF NON-BRANCHED *cata*-CONDENSED ARENES WITH $n = 4$ AND 5

Tetracene	(00) Tetracatafusene
Tetraphene	(01) Tetracatafusene
3.4-Benzophenanthrene	(11) Tetracatafusene
Chrysene	(12) Tetracatafusene
Pentacene	(000) Pentacatafusene
Picene	(121) Pentacatafusene
5.6-Benzochrysene	(112) Pentacatafusene
3:4,5.6-Dibenzophenanthrene	(111) Pentacatafusene
1:2,5:6-Dibenzanthracene	(102) Pentacatafusene
1:2,7:8-Dibenzanthracene	(101) Pentacatafusene
1:2-Benzotetraphene	(011) Pentacatafusene
3:4-Benzotetraphene	(012) Pentacatafusene
Pentaphene	(010) Pentacatafusene
1:2-Benzotetracene	(001) Pentacatafusene

REFERENCES

[1] Preceding part in the series, A. T. Balaban and F. Harary, *Rev. Roumaine Chim.* in press.

[2] Research supported in part by Grant GN-2544 from the National Science Foundation, USA, Chemical Information Program.

[3] E. Clar, *Polycyclic Hydrocarbons.* Academic Press–Springer Verlag, London (1964).

[4] *IUPAC, Nomenclature of Organic Chemistry,* 1957. Butterworths Scientific Publ., London (1958); *J. Am. Chem. Soc.* **82,** 5545 (1960).

[5] *The Naming and Indexing of Chemical Compounds from Chemical Abstracts.* Introduction to the Subject Index, *Chem. Abstr.* **56.** Jan–June (1962).

[6] A. M. Patterson, L. T. Capell and D. F Walker, *Ring Index* (2nd Edition). Am. Chem. Soc., Washington (1960).

[7] J. Lederberg, *Proc. Natl. Acad. Sci. U.S.A.* **53.** 134 (1965); NASA Reports CR-57029, CR-68898, CR-68899.

[8] B. Pullman, *The Application of Wave Mechanical Methods to the Study of Molecular Properties* (Edited by R. Daudel), *Adv. in Chem. Phys.* Vol. 8; p. 163. Interscience, London (1965).
G. M. Badger, *The Chemical Basis of Carcinogenic Activity* (Edited by I. N. Kugelmass). Ch. Thomas Publ., Springfield, Ill., Amer. Lecture Series (1962);
H. Busch, *An Introduction to the Biochemistry of the Cancer Cell* p. 207. Academic Press, New York (1962).

[9] A T. Balaban and C. D. Nenitzescu, *Friedel Crafts and Related Reactions* (Edited by G. A. Olah) Vol 2; p. 979. Wiley–Interscience (1964).

[10] A. T. Balaban, *Studii şi Cercetări Chem. Acad. R.S. România* **13,** 1299 (1965).

[11] *a* W. Jenny and R. Peter, *Angew Chem.* **77,** 1027 (1965); *Angew. Chem. internat. Ed.* **4,** 979 (1965);
b H. A. Staab and F. Binnig, *Chem. Ber* **100,** 293 (1967).

[12] M. Flammang-Barbieux, J. Nasielski and R. H Martin, *Tetrahedron Letters* 743 (1967).

[13] F. Harary, *Publ. Math. Inst. Hungar. Acad. Sci.* **5,** 63 (1960).

[14] F. Harary, *Applied Combinatorial Mathematics* (Edited by E. F Beckenbach) p. 185. Wiley, New York (1964).

[15] F. Harary, *Graph Theory and Theoretical Physics* (Edited by F. Harary) Chapter 1, Section 5. Academic Press, London (1967).

[16] F. Harary, *Ibid.* Chapter 1, Section 6.

[17] C A. Coulson and H. C. Longuet-Higgins, *Proc. Roy. Soc.* A **192,** 16 (1947).

[18] E. Clar, *Tetrahedron* **3,** 325 (1958); E. Clar and A. McCallum, *Ibid.* **10,** 171 (1960); E. Clar and I. A. McPherson, *Ibid.* **18,** 144 (1962); E. Clar, *Z. Chem.* **2,** 35 (1962); *Liebigs Ann. Chem.* **669,** 44 (1963); *Chimia* **18,** 375 (1964).

[19] V. E. Sahini, *J. Chim. Phys.* 177 (1962); *Rev. Chim. Acad. R.P. Roumaine* 7, 1265 (1962); *Rev. Chim. (Roumania)* 15, 551 (1964).

[20] M. Gordon and W. H. T. Davison, *J. Chem. Phys.* 20, 428 (1952).

[21] M. Saunders, *Tetrahedron* 23, 2105 (1967).