

Pólya⁶ also developed asymptotic formulas for other families of chemical compounds. For example, the simplest case of unsaturated hydrocarbons, $C_n H_{2n}$, containing exactly one double bond, are counted asymptotically by

$$\frac{\rho^{-n}}{4^n}$$

However, for more general unsaturated hydrocarbons $C_n H_{2n+2-2b}$, containing b double bonds, the asymptotic number is only known up to a constant:

$$K_7 = \rho^{-n} m^{(3b-5)/2},$$

where K_7 depends only on b .

Consider substituted paraffins with m distinct radicals substituted for hydrogen atoms, of the form $C_n H_{2n+2-m} X^m \dots X^{(m)}$. The asymptotic expression obtained by Pólya was

$$\rho^{-n} n^{(2m-5)/2} \frac{\rho a b^3}{4\sqrt{\pi}} \left(\frac{2}{\rho a b^2} \right)^2,$$

where ρ , a and b are the same constants as in the expressions for R_n and r_n .

Finally, Pólya also showed that the number of hydrocarbons with a given cyclic structure is asymptotically proportional to R_n , the number of alcohols, and that the constant of proportionality is determined by the cycle index of the group of the graph having the given cyclic structure.

REFERENCES

1. Pólya, G. (1937). Kombinatorische Anzahlbestimmung für Gruppen, Graphen und chemische Verbindungen, *Acta. Math.* 68, 145.
2. Harary, F. and Palmer, E. M., (1973). "Graphical Enumeration", Academic Press, New York, London and San Francisco.
3. Pólya, G., (1936). Algebraische Berechnung der Isomeren einiger organischer Verbindungen, *Z. Kristallogr. (A)*, 93, 414.
4. Pólya, G., (1935). Un problème combinatoire général sur les groupes des permutations et le calcul du nombre des composés organiques, *Compt. rend. Acad. Sci. Paris*, 201, 1167.
5. Pólya, G., (1936). Sur le nombre des isomères de certains composés chimiques, *Compt. rend. Acad. Sci. Paris*, 202, 1554.
6. Pólya, G., (1936). Über das Anwachsen des Isomerenzahlen in den homologen Reihen des organischen Chemie, *Vierteljahrsh. Naturforsch. Ges. Zürich*, 81, 243.
7. Cayley, A., (1874). On the mathematical theory of isomers, *Philos. Mag.* 67, 444.
8. Blair, C. M. and Henze, H. R., (1931). The number of structurally isomeric alcohols of the methanol series, *J. Amer. Chem. Soc.* 53, 3042.
9. Otter, R., (1948). The number of trees, *Ann. of Math.* 49, 583.
10. Robinson, R. W. and Schwenk, A. J. (1975). The distribution of degrees in a large random tree, *Discrete Math.* 12, 359.

A5955
A5962
A36672
A36673
A634
A635
A636
A642
A678
A598
A599
A600
A602
A631
A632
A625
A628

CHAPTER 4

The Enumeration of Acyclic Chemical Compounds*

Ronald C. Read

University of Waterloo,
Department of Combinatorics and Optimization,
Waterloo, Ontario, Canada

1. Introduction	25
2. Some straight-forward uses of the counting series for alkyl radicals	27
3. Disubstituted hydrocarbons	32
4. Unsubstituted alkanes	36
5. Stereo-isomers	40
6. General acyclic hydrocarbons — structural isomers	46
7. Enumeration of unlabelled stereohydrocarbons	53
References	102

1 INTRODUCTION

In the last chapter we saw one or two ways in which combinatorial methods, and, in particular, Pólya's Theorem, can be used to enumerate chemical compounds of various kinds. We shall now follow up this idea in greater detail.

The most general problem of this type would be the following. Given an empirical formula, which merely states how many atoms of each kind there are in a compound, determine the number of distinct chemical structures that have those numbers of atoms. For example, we might ask how many structural isomers there are having the empirical formula $C_{30}H_{51}N_7O_{15}$. More generally, we ask for a mathematical formula, or some automatic procedure, whereby the number of isomers corresponding to any given empirical formula can be

*The original work described in this chapter was carried out under grant number A8142 from the National Research Council of Canada.

determined. A different, and somewhat more realistic, version of this problem is to ask for the number of stereo-isomers corresponding to the given empirical formula. Unfortunately, both these problems are notoriously intractable,* and it is most unlikely that any practical solution to them will be found in the near future. However, if we lower our sights a little, and confine our attentions to some restricted class of chemical compounds, then a mathematical treatment of the problem becomes possible. In particular, the acyclic hydrocarbons form an important class of compounds for which an explicit enumeration is feasible, and it is to the counting of these compounds, and some of their derivatives, that we shall address ourselves in this chapter.

The enumeration of acyclic hydrocarbons containing at most one double or triple bond has been well studied in the past by Cayley, Blair and Henze, Pólya and others (see the references given below). Later, new techniques were discovered whereby these same enumerations could be obtained more easily, but which were not, apparently, used to derive any new results. This is surprising, since all the apparatus was then at hand for the solution of more difficult enumerative problems, including the one already alluded to of counting the general acyclic hydrocarbons, i.e. having any numbers of double or triple bonds. As it happened, this problem does not appear to have been tackled at all prior to the publication of my paper,⁷ the results of which will be extended here.

Thus the aim of this chapter will be threefold. First, it will present a short history of the early work in the counting of acyclic compounds; secondly, the more important of these enumerations will be derived, using (and thereby illustrating) the most up-to-date techniques for obtaining these results; and thirdly, the enumeration of both structural and stereo isomers of acyclic hydrocarbons in general (a hitherto neglected problem) will be given.

The compounds that we shall be enumerating fall naturally into families, within which they are ranked according to the number of carbon atoms that they have. This makes the generating function, or counting series, in which the power of x is the number of carbon atoms, the most suitable method of displaying our results. Our starting point will be the counting series that has already been derived in the previous chapter, namely the counting series $A(x)$ for the numbers, A_n , of structural isomers of the alkyl radicals, having the general formula $-C_nH_{2n+1}$. A word about notation. The symbols Q , R , X and Y will be used to denote unspecified atoms or radicals. When we need more symbols we shall use the letters Q and R with subscripts ($Q_1, Q_2, \dots, R_1, R_2, \dots$) even though, strictly speaking, this conflicts with the chemical usage of subscripts to indicate the number of atoms. This will be preferable to the use of a more complicated notation such as would be required to avoid this conflict. In any case, this conflict is more apparent than real, since the letters Q and R are not atom symbols.

*Substantial progress has recently been achieved by Masinter and his colleagues.⁸

2 SOME STRAIGHT-FORWARD USES OF THE COUNTING SERIES FOR ALKYL RADICALS

As we saw in Chapter 3, the recursion formula for this counting series $A(x) = A_0 + A_1x + A_2x^2 + \dots$ is $\int 2\phi$

$$A(x) = 1 + xZ(S_3; A(x)) \\ = 1 + \frac{1}{6}x[A^3(x) + 3A(x)A(x^2) + 2A(x^3)] \quad (2.1)$$

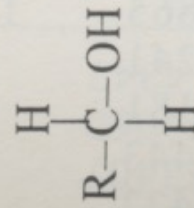
and from this equation the coefficients in $A(x)$ can be calculated to as many terms as one likes.

The calculation by hand of these coefficients by means of (2.1) is tedious, however, and their computation up to A_{20} by Blair and Henze¹ in 1931, and by Perry⁵ up to A_{30} in 1932, was a remarkable *tour de force*. A modern electronic computer can be persuaded to perform these calculations in a very short time, and the results of such a calculation are given in Table I. We shall now see how the numbers A_n and the series $A(x)$ can be used to enumerate several families of acyclic compounds.

We first note that if we join an alkyl radical to a hydroxyl group $-OH$ we get an alkyl alcohol of the methanol series $C_nH_{2n+1}OH$. The number of structural isomers of the alcohols with this formula is thus A_n . If $n=0$ we get the compound $H-OH$, which, of course, is not an alcohol in the chemical sense (or in any other sense!); but its inclusion in the series will do no harm, and will give some advantages.

Cayley³ and Blair and Henze¹ derived their general results by considering separately the number of primary, secondary and tertiary alcohols. We shall find it more convenient to derive these from the general result.

(a) Primary alcohols. The primary alcohols with n carbons are of the form



where R denotes an alkyl radical with $n-1$ carbon atoms. Hence there is a one-to-one correspondence between the primary alcohols with n carbons and the general alcohols with $n-1$ carbons. Thus the counting series for primary alcohols is

$$xA(x) \quad (2.2)$$

the multiplied x accounting for the difference of 1 in the number of carbon atoms. Since $H-OH$ is catered for in the counting series $A(x)$, the counting series $xA(x)$ takes account of the corresponding alcohol, methanol (CH_3OH). In a sense this is not a primary alcohol, but is best included here since it is certainly not a secondary or tertiary alcohol.

(b) Secondary alcohols. The secondary alcohols with n carbon atoms are of

The numbers of structural isomers of some acyclic compounds

<i>n</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
Alkyl radicals A(x)	1	1	2	4	8	17	39	89	211	507	1238	3057	7639
Primary alcohols	1	1	1	2	4	8	17	39	89	211	507	1238	3057
Secondary alcohols			1	1	3	6	15	33	82	194	482	1188	2988
Tertiary alcohols				1	1	3	7	17	40	102	249	631	1594
B(x)	1	1	2	3	7	14	32	72	171	405	989	2426	6045
Esters			1	2	4	9	20	45	105	249	599	1463	3614
Ethylene derivs.			1	1	3	5	13	27	66	153	377	914	2281
C _n H _{2n} XY	1	1	2	5	12	31	80	210	555	1479	3959	10652	28760
C _n H _{2n} X ₂	1	1	2	4	9	21	52	129	332	859	2261	5983	15976
Glycols			1	2	6	14	38	97	260	688	1856	4994	13550
P(x)	1	1	2	4	9	18	42	96	229	549	1347	3326	8330
Q(x) - A(x ²)	1	1	1	2	2	6	13	33	78	194	474	1188	2971
Alkanes	1	1	1	1	2	3	5	9	18	35	75	159	355
Alkyl radicals A(x)	19241	48865	124906	321198	830219	2156010	5622109	156010	830219	321198	830219	2156010	5622109
Primary alcohols	19241	48865	124906	321198	830219	2156010	5622109	156010	830219	321198	830219	2156010	5622109
Secondary alcohols	7528	19181	49060	126369	326863	849650	2216862	2156010	830219	326863	849650	2216862	5622109
Tertiary alcohols	4074	10443	26981	69923	182158	476141	1249237	2156010	830219	182158	476141	1249237	5622109
B(x)	15167	38422	97925	251275	648061	1679869	4372872	2156010	830219	648061	1679869	4372872	5622109
Esters	22695	57564	146985	377555	974924	2529308	6589734	2156010	830219	974924	2529308	6589734	5622109
Ethylene derivs.	14397	36564	93650	240916	623338	1619346	4224993	2156010	830219	623338	1619346	4224993	5622109
C _n H _{2n} XY	211624	576221	1572210	4297733	11767328	32266801	88594626	211624	576221	11767328	32266801	88594626	211624
C _n H _{2n} X ₂	115469	312246	847241	2304522	6283327	17164401	46972357	115469	312246	6283327	17164401	46972357	115469
Glycols	100302	273824	749316	2053247	5635266	15484532	42599485	100302	273824	749316	2053247	5635266	100302
P(x)	21000	53407	136639	351757	909962	2365146	6172068	21000	53407	136639	351757	909962	21000
Q(x) - A(x ²)	19142	49060	126280	326863	849439	2216862	5805749	19142	49060	126280	326863	849439	19142
Alkanes	1858	4347	10359	24894	60523	148284	366319	1858	4347	10359	24894	60523	1858
Alkyl radicals A(x)	14715813	38649152	101821927	269010485	712566567	269010485	712566567	14715813	38649152	101821927	269010485	712566567	14715813
Primary alcohols	5622109	14715813	38649152	101821927	269010485	712566567	269010485	5622109	14715813	38649152	101821927	269010485	5622109
Secondary alcohols	5806256	15256265	40210657	106273050	281593237	712566567	269010485	5806256	15256265	40210657	106273050	281593237	5806256
Tertiary alcohols	3287448	8677074	22962118	60915508	161962845	712566567	269010485	3287448	8677074	22962118	60915508	161962845	3287448
B(x)	11428365	29972078	78859809	208094977	550603722	712566567	269010485	11428365	29972078	78859809	208094977	550603722	11428365
Esters	17234114	45228343	119069228	314368027	832193902	712566567	269010485	17234114	45228343	119069228	314368027	832193902	17234114
Ethylene derivs.	11062046	29062341	76581151	202365823	536113477	712566567	269010485	11062046	29062341	76581151	202365823	536113477	11062046
C _n H _{2n} XY	243544919	670228623	1846283937	5090605118	14047668068	712566567	269010485	243544919	670228623	1846283937	5090605118	14047668068	243544919
C _n H _{2n} X ₂	128741107	353345434	970999198	2671347292	7356752678	712566567	269010485	128741107	353345434	970999198	2671347292	7356752678	128741107
Glycols	16166991	42488077	112004630	296080425	784688263	712566567	269010485	16166991	42488077	112004630	296080425	784688263	16166991
P(x)	15256265	40209419	106273050	281590180	747890675	712566567	269010485	15256265	40209419	106273050	281590180	747890675	15256265
Q(x) - A(x ²)	910726	2278658	5731580	14490245	36797588	712566567	269010485	910726	2278658	5731580	14490245	36797588	910726
Alkanes	910726	2278658	5731580	14490245	36797588	712566567	269010485	910726	2278658	5731580	14490245	36797588	910726

TABLE I

the form



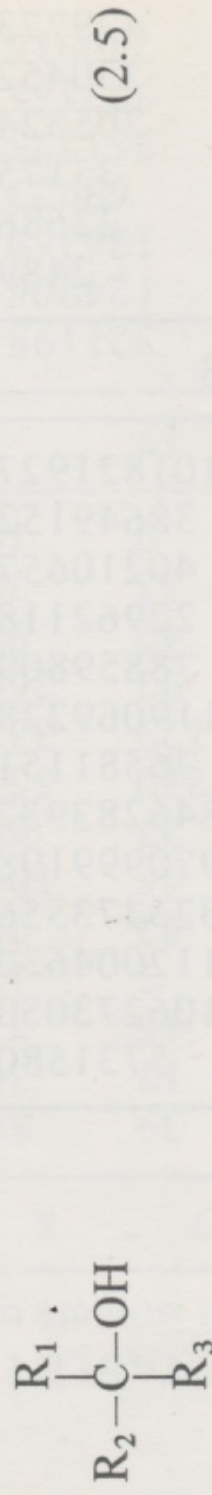
where R_1 and R_2 denote alkyl radicals that are not hydrogen atoms, having $n-1$ carbon atoms between them. To enumerate these isomers we apply Pólya's theorem. We have two boxes (R_1 and R_2) which can be interchanged (since we are counting *structural* isomers), and in each we must put an alkyl radical (not hydrogen). Hence the figure-counting series is $A(x)-1$, and the group is S_2 . Hence the numbers of ways of choosing R_1 and R_2 are given by the configuration-counting series

$$Z(S_2; A(x) - 1).$$

Taking into account the extra carbon atoms in (2.3) we see that the counting series for the secondary alcohols is

$$xZ(S_2; A(x) - 1) = \frac{1}{2}x[A^2(x) - 2A(x) + A(x^2)]. \quad (2.4)$$

(c) Tertiary alcohols. The tertiary alcohols have the formula



where R_1, R_2, R_3 are alkyl radicals (not hydrogen) having $n-1$ carbon atoms between them. Using Pólya's theorem with three boxes, permutable by S_3 , we obtain the counting series

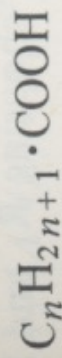
$$xZ(S_3; A(x) - 1) \\ \frac{1}{6}x[A^3(x) - 3A^2(x) + 3A(x)A(x^2) - 3A(x^2) + 2A(x^3)] \quad (2.6)$$

in much the same way as before. This reduces to

The reader can verify that; by adding the right-hand sides of (2.2), (2.4) and (2.6) we obtain the right-hand side of (2.1) except for the constant term which, corresponding to $H-OH$, is not included.

There are many other things that we can attach to an alkyl radical besides a hydroxyl group, and we see therefore that the counting series (2.1) will enumerate compounds of the form $C_nH_{2n+1}X$, i.e. monosubstituted saturated hydrocarbons, where X is a monovalent atom (other than hydrogen) or a monovalent radical (provided it can be clearly delineated from the rest of the molecule — it could not be an alkyl radical for example). If X contains carbon atoms then we need to juggle with the counting series slightly if we wish the

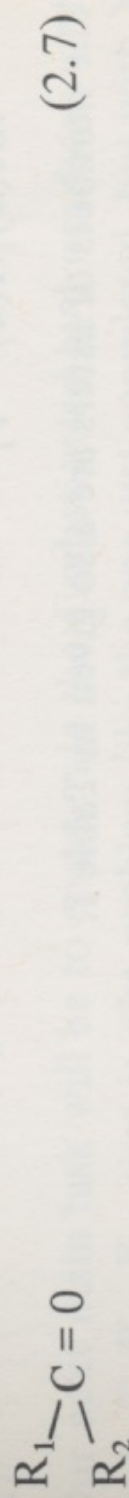
power of x to be the *total* number of carbon atoms in the molecule. Thus if X is the carboxyl group $-COOH$, the coefficient A_n will enumerate the isomers of



in which the total number of carbon atoms is $n+1$. The counting series for these acids by total number of carbon atoms is therefore $xA(x)$, in which A_n is the coefficient of x^{n+1} (Cf. the enumeration of primary alcohols).

The enumeration of compounds of the form $C_nH_{2n+1}X$ is therefore a trivial variation on that of enumerating the alcohols, and we quickly turn to some less trivial deductions from the basic result (2.1).

Aldehydes and ketones. It will be convenient to consider aldehydes and ketones together, to begin with, since then the general formula is



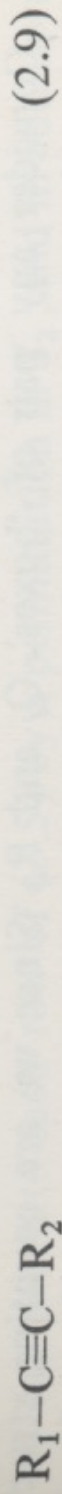
where R_1 and R_2 are alkyl radicals — possibly hydrogen atoms. We have here two interchangeable boxes in which to place a figure enumerated by $A(x)$. Hence, applying Pólya's theorem we obtain the configuration counting series $Z(S_2; A(x))$. This is similar to our treatment of secondary alcohols, and (as there) we must multiply by x to allow for the extra carbon in (2.7). Thus the counting series for aldehydes and ketones together is

$$B(x) = xZ(S_2; A(x)) \\ = \frac{1}{2}x[A^2(x) + A(x^2)] \quad A \text{ is } AS98 \quad (2.8)$$

The coefficients in $B(x)$ are tabulated in Table I.

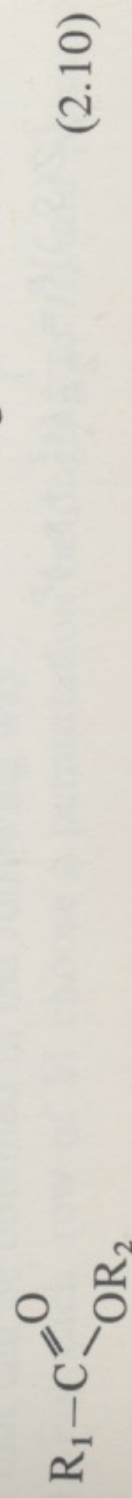
The numbers of aldehydes are easily found, since their general formula $C_nH_{2n+1} \cdot CHO$ is of the form $C_nH_{2n+1}X$. The counting series is therefore $xA(x)$, the same as, for example, that of the organic acids discussed above, which is hardly surprising considering the chemical affinity between these two types of compound. Subtracting this counting series from $B(x)$ we obtain the counting series for the ketones alone. The ketones are therefore the difference between $B(x)$ and the number of primary alcohols.

Another simple enumeration is that of the hydrocarbons of the acetylene series, having the general formula.



where R_1 and R_2 are alkyl radicals or hydrogen atoms. The combinatorial situation is the same as for aldehydes and ketones (two interchangeable boxes etc.) but there are now two extra carbon atoms. The counting series is therefore $xB(x)$, with $B(x)$ defined by (2.8).

A different situation arises if we consider the esters with general formula



where R_1 and R_2 are alkyl radicals, and where R_1 can be a hydrogen atom, whereas R_2 must not (since this would give an acid). If R_1 has r_1 carbon atoms and R_2 has r_2 carbon atoms, then the number of ways of choosing R_1 and R_2 in (2.10) to get a total of carbon atoms is

$$\sum A_{r_1} A_{r_2} \quad (2.11)$$

where the summation is for $r_1 + r_2 = n - 1$, $0 \leq r_1 \leq n - 2$ and $1 \leq r_2 \leq n - 1$. It is easily seen that (2.11) is the coefficient of x^n in

$$x(A_0 + A_1x + A_2x^2 + \dots)(A_1x + A_2x^2 + \dots).$$

Hence the counting series for these esters is

$$xA(x)[A(x) - 1] \quad (2.12)$$

The numbers of esters are also given in Table I.

3 DISUBSTITUTED HYDROCARBONS

The previous results were fairly immediate deductions from the basic result (2.1), using Pólya's Theorem. We now turn to the problem of enumerating the disubstituted hydrocarbons with the formula $C_nH_{2n}XY$; This can also be achieved using Pólya's theorem, but the method is less immediate and more difficult. We shall first assume that X and Y are *different* radicals or atoms.

Let us imagine that we pick up the molecule $C_nH_{2n}XY$ by the radicals X and Y , and pull them apart as far as we can. We see that we can write the structural formula as



where the Q_i and R_i are alkyl radicals (or hydrogen atoms). Concentrating on those molecules with a given value of p , we have the problem of placing alkyl radicals having a total carbon content of $n - p$ atoms, into $2p$ boxes. Thus we have an enumeration problem of the Pólya type; our remaining task is to discover the appropriate permutation group, and its cycle index.

Consider any pair of 'boxes' Q_i and R_j . Since we are counting structural isomers these boxes are interchangeable, i.e. can be permuted independently of the others, and it follows that the group we require is

$$S_2 \times S_2 \times \dots \times S_2 \quad (p \text{ factors})$$

whose cycle index is

$$\{Z(S_2)\}^p = \frac{1}{2^p} (s_1^2 + s_2)^p. \quad (3.2)$$

The number of ways of allocating alkyl radicals to these $2p$ boxes is therefore given by the coefficient of x^{n-p} in

$$\{\frac{1}{2}[A^2(x) + A(x^2)]\}^p$$

i.e. by the coefficient of x^n in

$$x^p \{ \frac{1}{2}[A^2(x) + A(x^2)] \}^p = B^p(x)$$

with $B(x)$ defined as in (2.8).

To find the total number of compounds of this type we must sum for $p = 1, 2, 3, \dots$. We obtain

$$\sum_{p=1}^{\infty} B^p(x) = \frac{B(x)}{1 - B(x)} \quad (3.3)$$

Our next main task will be to enumerate the disubstituted hydrocarbons with formula $C_nH_{2n}X_2$, that is, the problem like the one just completed, but with X and Y the same. To make the solution of this problem a little simpler, however, we shall digress to consider first the enumeration of the alkyl derivatives of ethylene.

These are the compounds having the general formula



where the R_i are alkyl radicals. From (3.4) we see that we have 4 boxes into which to put alkyl radicals - a straightforward Pólya problem provided we can determine the relevant permutation group.

We can interchange R_1 and R_2 , or R_3 with R_4 , or both of these, and have the same compound. In addition, we can turn the diagram (3.4) upside-down. If we think of R_1, R_2, R_3 and R_4 as elements of a matrix

$$\begin{pmatrix} R_1 & R_2 \\ R_3 & R_4 \end{pmatrix} \quad (3.5)$$

then we can permute each row of (3.5) by a permutation in S_2 ; and in addition we can permute the two rows by a permutation in S_2 . The resulting permutation group is a special case of what Pólya called the 'wreath product' of two groups, a concept that we must briefly discuss.

Let M be an $m \times n$ matrix, and let G and H be two permutation groups of degrees m and n respectively. Now consider all permutations of the mn elements of M that can be obtained in the following way:

(a) For each row of M choose a permutation belonging to H , and permute

the elements of that row by that permutation. The permutations for the various rows are chosen independently; they need not be all the same, nor need they all be different.

(b) Having permuted each row according to the permutation chosen for it, now permute the rows among themselves by a permutation belonging to G .

The set of all permutations of the mn elements of M that are obtainable in this way form a group. It is called the 'wreath product' of G and H and is denoted by $G[H]$.

Pólya,⁶ gave a formula for the cycle index of a wreath product in terms of those of the two groups. Suppose, for brevity, we write

$$Z(G) = g(s_1, s_2, s_3, \dots, s_m)$$

and

$$Z(H) = h(s_1, s_2, s_3, \dots, s_n)$$

Then the cyclic index of the wreath product $G[H]$ is obtained by replacing each occurrence of s_i in $g(s_1, s_2, s_3, \dots, s_m)$ by $h(s_i, s_{2i}, s_{3i}, \dots, s_{mi})$. In the original notation for cycle indices this can be written as

$$Z(G[H]) = Z(G; Z(H; s_1, s_2, s_3, \dots, s_m), Z(H; s_2, s_4, s_6, \dots, s_{2m}), \dots, Z(H; s_n, s_{2n}, s_{3n}, \dots, s_{mn})) \quad (3.6)$$

It will be seen that the group of permutations of the elements R_1, R_2, R_3, R_4 of the matrix (3.5) is obtained in just this way, taking G and H to be S_2 . Hence it is the wreath product $S_2[S_2]$. We could easily have found the cycle index of this group by simply listing its 8 permutations, but we shall make use of (3.6) instead. This is akin to using a steam-hammer to crack a nut, but it will serve to demonstrate the mechanism of the hammer.

In $Z(G) = \frac{1}{2}(s_1^2 + s_2)$ we have to replace each occurrence of s_1 by $Z(H) = \frac{1}{2}(s_1^2 + s_2)$, and each occurrence of s_2 by $\frac{1}{2}(s_2^2 + s_4)$. We then have

$$\begin{aligned} Z(S_2[S_2]) &= \frac{1}{2}\{[\frac{1}{2}(s_1^2 + s_2)]^2 + [\frac{1}{2}(s_2^2 + s_4)]\} \\ &= \frac{1}{8}\{s_1^4 + 2s_1^2 + 3s_2^2 + 2s_4\}. \end{aligned}$$

We can now resume the enumeration that we interrupted in order to define the wreath product. We have the appropriate cycle index, and since the figure-counting series is $A(x)$ we derive the configuration-counting series

$$\frac{1}{8}\{A^4(x) + 2A^2(x)A(x^2) + 3A^2(x^2) + 2A(x^4)\}. \quad (3.8)$$

This is the first example where we have used the wreath product, but we shall need it again for the enumeration of the compounds $C_nH_{2n}X_2$ to which we now return.

Let us first suppose that the number p of carbon atoms between the two X 's

is an even number, $p = 2k$ say. Then we can write the general formula as



much as before (3.1). This time, however, the group will be different, because we can now reverse the diagram (3.9) end-for-end and have an equivalent scheme for the molecule. The group of permutations for the alkyl radicals in the left-hand half of the molecules (i.e. Q_i and R_i for $i = 1, 2, \dots, k$) is

$$S_2 \times S_2 \times \dots \times S_2 \quad (k \text{ factors})$$

[Compare (3.2)]. The same is true for the right-hand half. Since the two sides can be permuted, the group for the whole diagram is the wreath product

$$S_2[S_2 \times S_2 \times \dots \times S_2] \quad (3.10)$$

To see the analogy with the ethylene example, and with the definition of wreath product, think of the $2k$ boxes on the left as forming the first row of a matrix while those on the right form the second row. Then each row can be permuted by $S_2 \times S_2 \times \dots \times S_2$; while the rows can be permuted bodily by S_2 . The cycle index of (3.10) is

$$\frac{1}{2}\{[\frac{1}{2}(s_1^2 + s_2)]^{2k} + [\frac{1}{2}(s_2^2 + s_4)]^k\}$$

and on substituting the figure-counting series $A(x)$, and multiplying by x^{2k} to accommodate the $2k$ carbon atoms between the X 's, we obtain the series

$$\begin{aligned} \frac{1}{2}\left\{\frac{x}{2}(A^2(x) + A(x^2))^{2k} + \left[\frac{x^2}{2}(A^2(x^2) + A(x^4))\right]^k\right\} \\ = \frac{1}{2}\{B^{2k}(x) + B^k(x^2)\}. \end{aligned} \quad (3.11)$$

When p is odd, say $p = 2k + 1$, then the above remarks apply to all the 'boxes' except the two attached to the middle carbon atom of the chain. These latter can be permuted among themselves independently of what happens to the others. Hence our group is now

$$S_2 \times S_2[S_2 \times S_2 \times \dots \times S_2]$$

which eventually gives us the counting series

$$\frac{1}{2}B(x)\{B^{2k}(x) + B^k(x^2)\}, \quad (3.12)$$

as the analogue of (3.11).

We now sum (3.11) and (3.12) over all values of k . Since we need $k = 0$ when p is odd (to include CH_2X_2) we shall include it also for p even. This will introduce a constant term, corresponding to the compound $X-X$. This inclusion

does no harm, and simplifies the result. We have

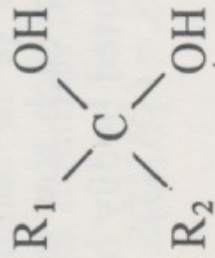
$$\frac{1}{2} \sum_{k=0}^{\infty} \{B^{2k}(x) + B^k(x^2)\} + \frac{1}{2} \sum_{k=0}^{\infty} B(x) \{B^{2k}(x) + B^k(x^2)\}$$

which reduces to

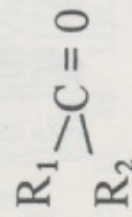
$$\frac{1}{2} \cdot \frac{1}{1-B(x)} + \frac{1}{2} \cdot \frac{1+B(x)}{1-B(x^2)} \quad (3.13)$$

By substituting specific radicals or atoms (Cl, OH, COOH etc.) for X and Y in the preceding enumerations, we can obtain the numbers of compounds of various specific types. One important family of compounds that can *almost* be enumerated in this way is that of the glycols. These are of the form $C_2H_{2n}(OH)_2$, i.e. they are of the form $C_2H_{2n}X_2$ just considered, except that the two hydroxyl groups must not be attached to the same carbon atom.

Hence to enumerate the glycols we must subtract from the numbers given by (3.13) the numbers of 'invalid' glycols. These will have the general formula



and are therefore equinumerous with the compounds



(aldehydes and ketones) which have already been enumerated (2.8). Thus the numbers of glycols can be found. They are given in Table I, along with the numbers of the two kinds of disubstituted hydrocarbons that we have just considered. The numbers of ethylene derivatives, first found by Blair and Henze,² are also included.

4 UNSUBSTITUTED ALKANES

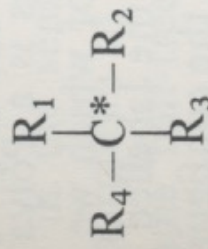
It might be thought that the enumeration of the alkanes, C_nH_{2n+2} , would present less of a problem than those considered hitherto, in so far as the alkanes are, so to speak, the compounds from which the others that we have considered can be derived; but this is not the case. The reason is that, for example, the X in $C_nH_{2n+1}X$ provides a convenient distinguishable point at which to attack the problem, while for the alkanes there is no such distinguished atom to pick up the problem by. In graph-theoretical terms the previous enumerations have been of

trees that were *rooted* in some way; what we shall now consider is a problem in the enumeration of unrooted trees.

The early researchers in this field (Cayley, Blair and Henze, and even Pólya) had a great deal of difficulty with this problem, and were forced to solve it by complex methods that made use of the centre and bicentre of the tree structure given by the structural formula. More recent developments have provided a powerful tool by which the enumeration of unrooted trees can be performed with comparative ease. We shall examine this tool shortly, but first we need some preliminary results.

Some of the advantages of having a carbon atom at which a substitution has taken place can be regained if, without substituting for any of the hydrogen atoms in an alkane molecule, we contrive to distinguish one carbon atom from all the rest. This corresponds to a feasible chemical procedure — the labelling of a molecule by replacing one carbon atom by a radio-isotope. We shall call such a molecule an 'alkane with a labelled carbon atom', and we shall denote the labelled carbon atom by C^* .

We shall now enumerate these labelled alkanes, whose general formula is



the R_i 's being alkyl radicals. Since we are considering structural isomers only, the four 'boxes' into which these radicals go can be permuted in any way, i.e. by any permutation of S_4 .

We therefore apply Pólya's theorem with $A(x)$ as the figure-counting series, and S_4 as the group. Since the cycle index of the group is

$$\frac{1}{24} (s_1^4 + 6s_1^2s_2 + 3s_2^2 + 8s_1s_3 + 6s_4)$$

and since we must multiply by x to accommodate the extra (labelled) carbon atom, our configuration counting series takes the form

$$\begin{aligned} P(x) &= \sum_{n=1}^{\infty} P_n x^n \\ &= xZ(S_4; A(x)) \end{aligned}$$

$$= \frac{1}{24} x \{A^4(x) + 6A^2(x)A(x^2) + 3A^2(x^2) + 8A(x)A(x^3) + 6A(x^4)\} \quad (4.1)$$

where P_n is the number of these compounds with n carbon atoms in all.

The other preliminary result that we require is the enumeration of alkanes in which one carbon-to-carbon valency bond has been labelled, that is, distin-

A602

A678

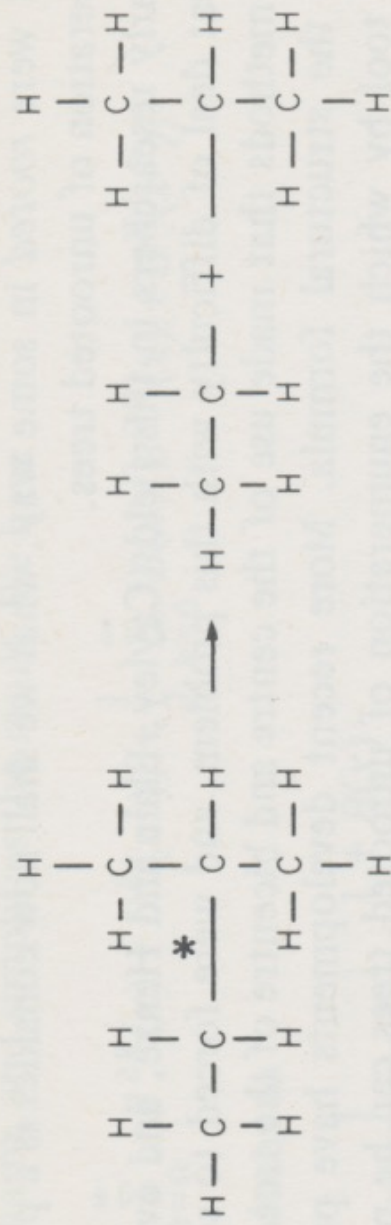


FIG. 1

guished from the other bonds. This labelling does not correspond in any direct way to a feasible chemical procedure, but is only a means to an end. As with the labelling of a carbon atom, this labelling of a bond gives us a point to start from. If we 'break' the labelled bond in two, we get two alkyl radicals as shown in Fig. 1 (in which the labelled bond is indicated by the asterisk).

Conversely, by joining two alkyl radicals, in an obvious way, we obtain an alkane in which one carbon-to-carbon bond is distinguished from the rest by virtue of being the common 'free bond' of the two radicals.

Hence the enumeration of these 'bond-labelled' molecules is a Pólya type problem. There are two boxes, permutable by S_2 , into each of which we must place an alkyl radical, but not just a hydrogen atom (since the labelled bond must be a carbon-to-carbon bond). Pólya's theorem then gives the configuration counting series

$$\begin{aligned}
 Q(x) &= \sum_{n=1}^{\infty} Q_n x^n \\
 &= Z(S_2; A(x) - 1) \\
 &= \frac{1}{2} \{ (A(x) - 1)^2 + A(x^2) - 1 \}
 \end{aligned} \tag{4.2}$$

We now look at the comparatively recent advance (alluded to above) which enables us to derive the counting series for unsubstituted (and unlabelled) alkanes from the two series $P(x)$ and $Q(x)$ just given (see the paper by Harary and Norman⁴). We shall not give the proof here, but discuss the theorem in the context of trees in general.

Let T be any tree. There will be certain one-to-one mappings (if only the identity) of the vertex set of T onto itself which leave the tree invariant, i.e. adjacent vertices remain adjacent, etc. These mappings (called automorphisms) form a group, the automorphism group of T . If there is an automorphism which maps a vertex u onto a vertex v we say that u and v are equivalent. This relation between vertices is an equivalence relation, as its name suggests, and divides the vertex set of T into equivalence classes. We shall let p^* denote the number of equivalence classes.

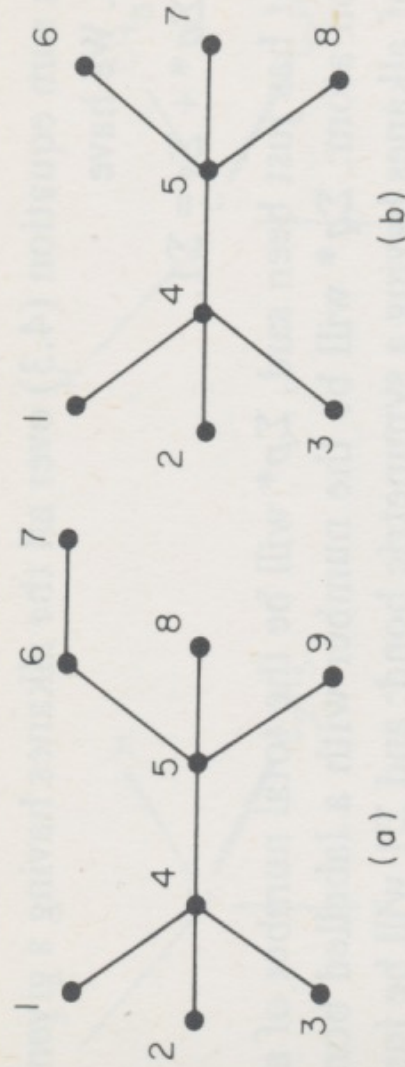


FIG. 2

Clearly any automorphism induces, in an obvious way, a mapping of the set of edges of T onto itself. If an edge e maps onto an edge f under some automorphism of T we say e and f are equivalent, and in this way we define equivalence classes of edges. Let q^* be the number of equivalence classes. The result that we shall use, and which is given by Harary and Norman⁴ is that

$$p^* - q^* + s = 1 \tag{4.3}$$

Here $s = 1$ if the tree has a symmetric edge (i.e. an edge uv which maps onto itself, in the sense that u and v map onto v and u respectively, under some automorphism of T), and $s = 0$ otherwise. A symmetric edge must clearly have the same 'half-tree' at its two ends, and hence there can be at most one such edge in any tree. The following two examples should make this clear.

In the tree of Fig. 2(a) vertices 1, 2, 3 are equivalent, and so are 8 and 9. All other vertices are in classes by themselves. Hence $p^* = 6$. The edges (1, 4), (2, 4) and (3, 4) are equivalent; so are (5, 8) and (5, 9). Hence $q^* = 5$. There is no symmetric edge, so $s = 0$ and (4.3) is satisfied. In Fig. 2(b) vertices 1, 2, 3, 6, 7, 8 are all equivalent; so are 4 and 5. Hence $p^* = 2$. As for the edges, edge (4, 5) is a symmetric edge and is in a class by itself. All other edges make up a second equivalence class. Hence $q^* = 2$. Since now $s = 1$, equation (4.3) is again satisfied.

This result is true for the alkanes, which are trees in the usual graph-theoretical sense if we ignore the hydrogen atoms. The vertices are then the carbon atoms and the edges are the carbon-to-carbon bonds. We do not alter the problem by ignoring the hydrogen atoms since they can be uniquely restored when the rest of the molecule is known. Suppose we take two replicas of an alkane molecule and label a carbon atom in each. If the two carbon atoms so labelled are equivalent vertices then the two labelled molecules are indistinguishable; if they are in different equivalence classes then we shall have two differently labelled molecules. Hence the number of *different* labelled molecules that can be obtained by labelling one carbon atom in an unlabelled alkane is p^* , the number of vertex-equivalence classes. In an exactly similar way we show that by labelling a carbon-to-carbon bond in an alkane we can get exactly q^* different alkanes with a labelled bond.

Now let us sum equation (4.3) over all the alkanes having a given number n of carbon atoms. We have

$$\Sigma p^* - \Sigma q^* + \Sigma s = \Sigma 1. \quad (4.5)$$

Now, by what has just been said, Σp^* will be the total number of alkanes with a labelled carbon atom; Σq^* will be the number with a labelled bond; Σs will be the number of alkanes having a symmetric bond; and $\Sigma 1$ will be just the number of alkanes — the number we are looking for. Thus Σp^* is the number P_n given by (4.1) and Σq^* is the number Q_n given by (4.2). If we can determine the number Σs of alkanes with a symmetric bond we shall know everything in (4.5) except the number we wish to find.

A molecule with a symmetric bond must have an even number (say $2k$) of carbon atoms, and, as already remarked, the 'half-trees' at the end of the symmetric bond must be the same. Hence, if we split the symmetric bond (as in the determination of Q_n) we get two alkyl radicals as before, but now they are identical. Thus, having chosen one of the A_k alkyl radicals for one half, we have determined the whole molecule. Hence the number of alkanes having a symmetric edge is $A_{n/2}$ (which we interpret to be 0 if n is odd). Denoting the required number of alkanes by C_n we thus rewrite (4.5) as

$$P_n - Q_n + A_{n/2} = C_n. \quad (4.6)$$

If we now multiply (4.6) by x^n , sum from $n = 1$ upwards, and rearrange, we have

$$\begin{aligned} C(x) &= \sum_{n=1}^{\infty} C_n x^n \\ &= P(x) - Q(x) + A(x^2) - 1 \end{aligned} \quad (4.7)$$

Since $P(x)$, $Q(x)$ and $A(x)$ are known we can calculate the series $C(x)$ and thus determine the number of alkanes. They are given in the last row of Table I. If we write $A_1(x)$ for $A(x) - 1$, then (4.7) can be put in the more elegant form

$$C(x) = P(x) - \frac{1}{2}\{A_1^2(x) - A_1(x^2)\}. \quad (4.8)$$

5 STEREO-ISOMERS

So far we have confined ourselves to problems of counting the *structural* isomers of the compounds that we have considered. This is chemically rather unrealistic, the enumeration of *stereo-isomers* being rather more to the point. However, both problems are of interest, and once one has been solved, the solution of the other follows readily. It therefore makes little difference which we tackle first.

We start by considering the monosubstituted alkanes (or alkyl radicals) as before, and we observe that the simplest such compound for which there are

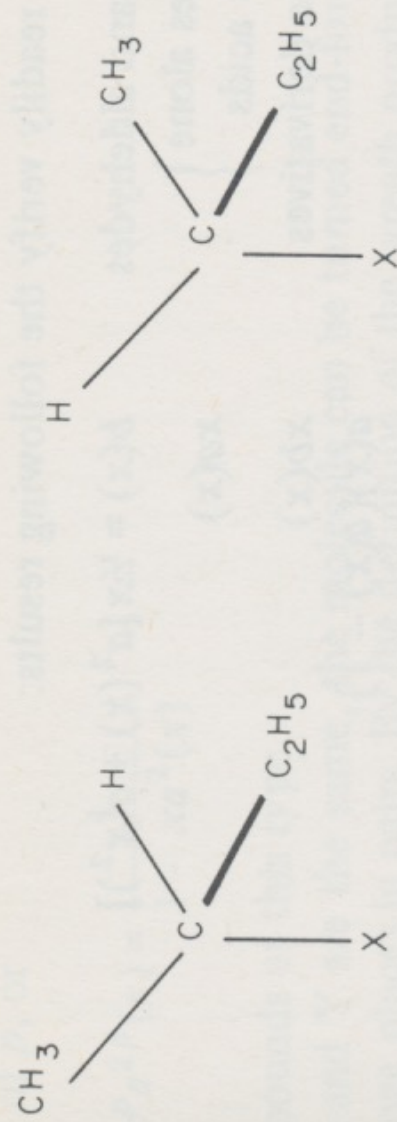
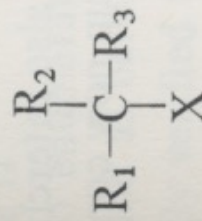


FIG. 3

stereo-isomers has the formula C_4H_9X , and can exist in the two enantiomorphous forms given in Fig. 3.

The X, H, CH_3 and C_2H_5 are at the vertices of a regular tetrahedron, and in both drawings in Fig. 3 the ethyl radical is to be thought of as above the plane of the paper. At once we see the difference that stereo-isomerism produces. As before, our general formula is



but instead of our being able to allow *any* permutation of the three 'boxes' into which R_1 , R_2 and R_3 are to be put, we can allow only those corresponding to a cyclic permutation.

Thus the problem is basically the same as before; only the group is different. We shall use lower-case letters in the stereo-isomer problems to correspond to the upper-case letters that we used in the enumeration of structural isomers, so that, for the monosubstituted alkanes, we shall write the generating function as

$$a(x) = a_0 + a_1x + a_2x^2 + \dots \quad (5.1)$$

The equation analogous to (2.1) is then

$$\begin{aligned} a(x) &= 1 + xZ(C_3; a(x)) \\ &= 1 + \frac{1}{3}x\{a^3(x) + 2a(x^3)\} \end{aligned} \quad (5.2)$$

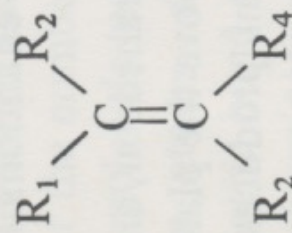
since $Z(S_3) = \frac{1}{3}(s_1^3 + 2s_3)$. This enumerates for example, the stereo-isomers of the alcohols of the methanol series. Primary alcohols are enumerated exactly as before. Secondary alcohols have the form (2.3), but now the radicals R_1 and R_2 are not interchangeable. Tertiary alcohols have the form (2.5), but R_1 , R_2 and R_3 may only be permuted cyclically. From these remarks we deduce the following counting series:

Primary alcohols	$xa(x)$
Secondary alcohols	$\frac{1}{2}x[a(x) - 1]^2$
Tertiary alcohols	$\frac{1}{3}x\{[a(x) - 1]^3 + 2[a(x^3) - 1]\}$.

We can also readily verify the following results:

Ketones and aldehydes	$b(x) = \frac{1}{2}x[a^2(x) + a(x^2)]$
Aldehydes alone	$xa(x)$
Alkanoic acids	$xb(x)$
Acetylene derivatives	$a(x)\{a(x) - 1\}$
Esters	

The enumeration of the stereo-isomers of the ethylene derivatives is somewhat different. The basic formula is



as before, but now we cannot interchange R_1 and R_2 unless we also interchange R_3 and R_4 . Bearing this in mind we see that the appropriate group now contains only the following permutations (with an obvious notation)

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$$

and its cycle index is therefore $\frac{1}{4}(s_1^4 + 3s_2^2)$. Hence the counting series for ethylene derivatives is

$$\frac{1}{4}x^2 [a^4(x) + 3a^2(x^2)]. \quad (5.3)$$

Turning now to the disubstituted alkanes we see that when $X \neq Y$ we can draw the molecule in a standard way, so as to make X and Y and the chain of p carbon atoms joining them lie in the plane of the paper, as in Fig. 4.

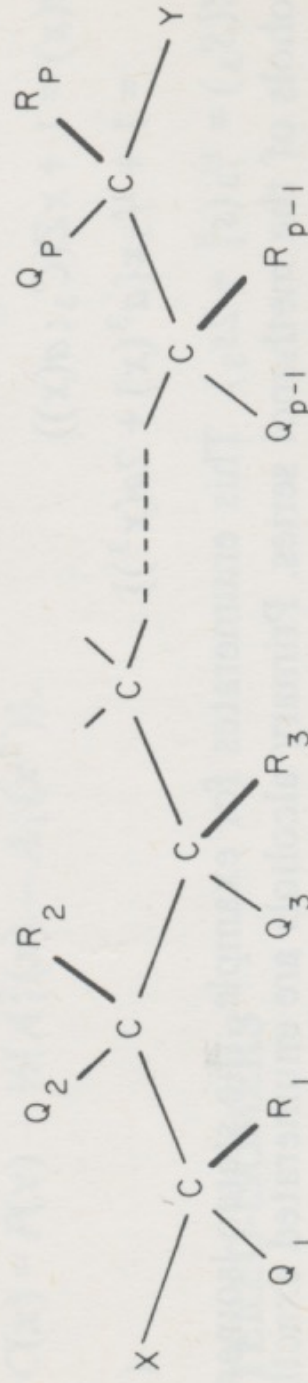


FIG. 4

(The bonds drawn with a heavy line in Fig. 4 are those which project above the plane of the paper.) We see that no permutation of the alkyl radicals Q_i and R_i is possible. Hence the group for this problem is the identity group E_{2p} of degree $2p$, whose cycle index is s_1^{2p} . The required counting series is thus $x^p a^{2p}(x)$ for a

given value of p , or

$$\sum_{p=0}^{\infty} x^p a^{2p}(x) = \frac{1}{1 - xa^2(x)}$$

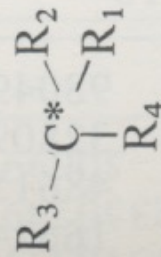
for all compounds of this type.

When X and Y are the same, the molecule can be turned end-for-end and the 'boxes' change places in pairs. By the definition of the wreath product the group in question is now $S_2 [E_p]$. (Note: this is irrespective of the parity of p , since if p is odd the 'boxes' attached to the central carbon atom in the chain change places along with the others.) The cycle index of this group is $\frac{1}{2}(s_1^{2p} + s_2^p)$, and hence the counting series for all compounds of this type is

$$\sum_{p=0}^{\infty} \frac{x^p}{2} \{a^{2p}(x) + a^p(x^2)\} = \frac{1}{2} \left\{ \frac{1}{1 - xa^2(x)} + \frac{1}{1 - xa(x^2)} \right\}$$

The numbers of the compounds considered so far in this section are tabulated in Table II.

To enumerate the alkanes, having regard to stereo-isomers, we use the same method as before, and first enumerate alkanes with a labelled carbon atom. It is readily verified that the 'boxes' in the diagram



can be permuted by any permutation of the alternating group A_4 , whose cycle index is

$$\frac{1}{12} (s_1^4 + 3s_2^2 + 8s_1s_3)$$

Applying Pólya's theorem, with $a(x)$ as the figure-counting series, we obtain

$$p(x) = \sum_{n=1}^{\infty} \frac{1}{12} x \{a^4(x) + 3a^2(x^2) + 8a(x)a(x^3)\} \quad (5.5)$$

as the stereo-analogue of (4.1).

The counting series for alkanes in which a carbon-carbon bond has been labelled is

$$\begin{aligned} q(x) &= Z(S_2; a(x) - 1) \\ &= \frac{1}{2} \{ [a(x) - 1]^2 + a(x^2) - 1 \}, \end{aligned} \quad (5.6)$$

the analogue of (4.2).

We now use the theorem, described and used in Section 4, to the effect that $p^* - q^* + s = 1$, and obtain the result

$$c(x) = p(x) - q(x) + a(x^2) - 1 \quad (5.7)$$

The numbers of stereo-isomers of some acyclic compounds

		1	2	3	4	5	6	7	8	9	10	11	12	
Alkyl radicals A(x)	Alkyl radicals A(x)	1	1	2	5	11	28	74	199	551	1553	4436	12832	
	Primary alcohols	1	1	1	2	5	11	28	74	199	551	1553	4436	
	Secondary alcohols			1	2	5	14	36	98	273	768	2197	6360	
	Tertiary alcohols				1	1	3	10	27	79	234	686	2036	
	b(x)	1	1	2	3	8	18	47	123	338	935	2657	7616	
	Esters		1	2	4	10	25	64	172	472	1319	3750	10796	
	Ethylene derivs.		1	1	4	6	18	42	118	314	895	2521	7307	
	C ⁿ H ^m XY	1	3	8	23	69	208	636	1963	6099	19059	59836	188576	
	C ⁿ H ^m X ₂	1	2	5	13	37	108	325	993	3070	9564	29979	94392	
	p(x)	1	1	2	10	22	60	158	439	1229	3525	10178	29802	
	Alkanes	1	1	1	2	5	11	24	55	136	345	900		
	Alkyl radicals A(x)	Alkyl radicals A(x)	37496	110500	328092	980491	328092	110500	37496	110500	328092	980491	2946889	8901891
Primary alcohols		12832	37496	110500	328092	980491	328092	110500	37496	110500	328092	980491	2946889	8901891
Secondary alcohols		18584	54780	162672	486154	1461197	4413988	13393855	4716540	1541014	5153883	15599094	22295746	
Tertiary alcohols		6080	18224	54920	166245	505201	1541014	4716540	15599094	5153883	15599094	22295746		
b(x)		22138	64886	191873	571169	1711189	5153883	15599094	22295746	64886	191873	571169	1711189	5153883
Esters		31416	92276	273172	814246	2441688	7360877	22295746						
Ethylene derivs.		21238	62566	185310	553288	1660490	5011299	15190665						
C ⁿ H ^m XY		596252	1890548	6008908	19139155	61074583	195217253	624913284						
C ⁿ H ^m X ₂		298311	945592	3005021	9570559	30539044	97611676	312462096						
p(x)		87862	261204	781198	2350249	7105081	21577415	65787902						
Alkanes		2412	6563	18127	50699	143255	408429	1173770						
Alkyl radicals A(x)		Alkyl radicals A(x)	82300275	251670563	772160922	2376294040	7333282754	22688455980						
	Primary alcohols	27012286	82300275	251670563	772160922	2376294040	7333282754							
	Secondary alcohols	40807290	124783669	382842018	1178140280	3635626680	11247841224							
	Tertiary alcohols	14480699	44586619	137648341	425992838	1321362034	4107332002							
	b(x)	47415931	144692886	443091572	1361233280	4194107380	12957209782							
	Esters	67819576	207083944	634512581	1950301202	6011920720	18581123978							
	Ethylene derivs.	46244031	141296042	433204573	1332261200	4108833222	12704949506							
	C ⁿ H ^m XY	2003090071	6428430129	20653101216	66420162952	213802390264	688796847976							
	C ⁿ H ^m X ₂	1001554565	3214232129	10326580526	33210135104	106901289420	344398593149							
	p(x)	201313311	618040002	1903102730	5876174472	18189503139	56435742554							
	Alkanes	3396844	9892302	28972080	85289390	252260276	749329719							

TABLE II

三

47

where $c(x) = \sum_{n=1}^{\infty} c_n x^n$, and c_n is the number of stereo-isomers with the formula $C_n H_{2n+2}$. This is the analogue of (4.7). The values of c_n are tabulated in the last row of Table II.

6 GENERAL ACYCLIC HYDROCARBONS—STRUCTURAL ISOMERS

The results derived so far in this chapter have all been results that have been known for some time; though, as already remarked, the methods used to obtain them here were not usually those by which they were first derived. We now apply the methods and techniques of this chapter to a new and more general enumeration — one that will contain many of the previous results as special cases. We shall study the enumeration of acyclic hydrocarbons having a given number of carbon atoms and given numbers of double and triple bonds, an enumeration that has been briefly described by myself.⁷ As before, we start by looking at the enumeration of monosubstituted hydrocarbons or, what amounts to the same thing, monovalent hydrocarbon radicals. This problem, though used here as a means to an end, is of some interest in its own right.

Let $G_{n,d,t}$ be the number of structural isomers of the monovalent acyclic radicals having n carbon atoms, d double bonds and t triple bonds. The formula of such a radical is $C_n H_{2n+1-2d-4t}$. We shall obtain a counting series of the form

$$G(x, y, z) = \sum_{n=0}^{\infty} \sum_{d=0}^{\infty} \sum_{t=0}^{\infty} G_{n,d,t} x^n y^d z^t. \quad (6.1)$$

Now such a radical, occurring as part of a molecule, will be connected to the rest of the molecule by a single bond. We shall need to consider also the number of divalent radicals connected to the rest of the molecule by a double bond — let the number having parameters n , d and t defined as above be $H_{n,d,t}$ — and the numbers of trivalent radicals connected to the rest of the molecule by a triple bond — let the corresponding number be $I_{n,d,t}$. Diagrammatically they are as shown in Fig. 5.

Thus $G_{n,d,t}$ is the number of radicals of the form shown in Fig. 5(a); $H_{n,d,t}$ is the number of the form shown in Fig. 5(b), and $I_{n,d,t}$ is the number of the form shown in Fig. 5(c). We shall call these three types of radicals 'G-radicals', 'H-radicals', and 'I-radicals' respectively. The shaded 'balloons' here represent the rest of the radical. Note that the double and triple bonds of attachment to the rest of the molecule are included in the count for d and t .

In conformity with (6.1) we define

$$H(x, y, z) = \sum_{n=0}^{\infty} \sum_{d=0}^{\infty} \sum_{t=0}^{\infty} H_{n,d,t} x^n y^d z^t \quad (6.2)$$

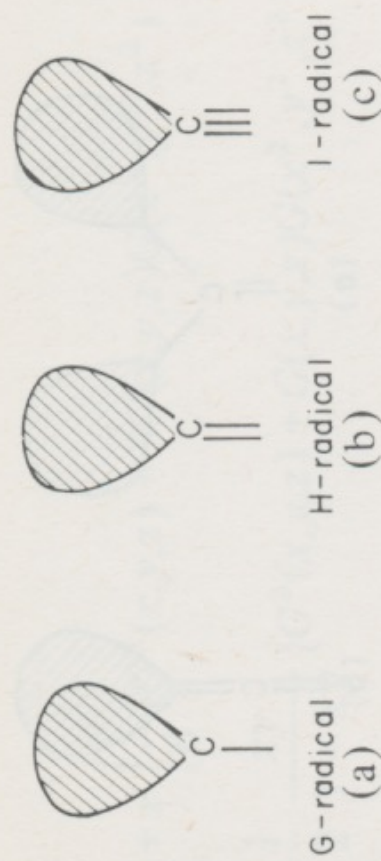


FIG. 5

$$I(x, y, z) = \sum_{n=0}^{\infty} \sum_{d=0}^{\infty} \sum_{t=0}^{\infty} I_{n,d,t} x^n y^d z^t \quad (6.3)$$

When examined in more detail, the radicals enumerated by $G(x, y, z)$ are seen to be of the three types shown in Fig. 6.

In Fig. 6(a) we have three boxes, permutable among themselves, into each of which we must put a G-radical. Since a box can contain a hydrogen atom alone we shall regard a hydrogen atom as a G-radical. The numbers of ways of forming G-radicals of this type are therefore enumerated by the series

$$xZ(S_3; G(x, y, z))$$

by Pólya's theorem. Note that there is a multiplied x since an extra carbon atom is present over and above the carbon content of the radicals in the boxes.

In Fig. 6(b) we have two dissimilar boxes. In one we must put a G-radical while in the other we must put an H-radical. Having regard to the extra carbon atom we see that the G-radicals of this type are enumerated by

$$xG(x, y, z)H(x, y, z).$$

In Fig. 6(c) we have only one box, in which to place an I-radical. The corresponding counting series here is therefore just $xI(x, y, z)$.

Putting these results together, and adding a constant term (= 1) to represent the G-radical consisting of a single hydrogen atom (which would not otherwise be included) we obtain the following

$$G(x, y, z) = 1 + x \{ Z(S_3; G(x, y, z)) + G(x, y, z)H(x, y, z) + I(x, y, z) \} \quad (6.4)$$

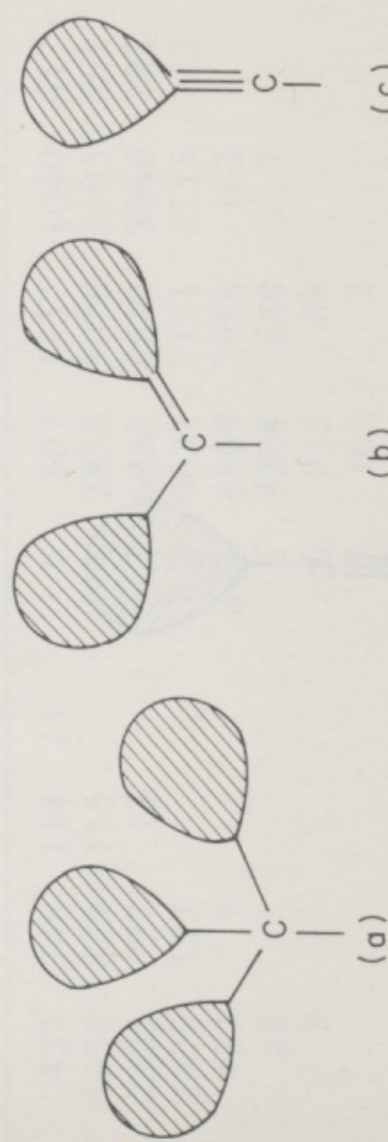


FIG. 6

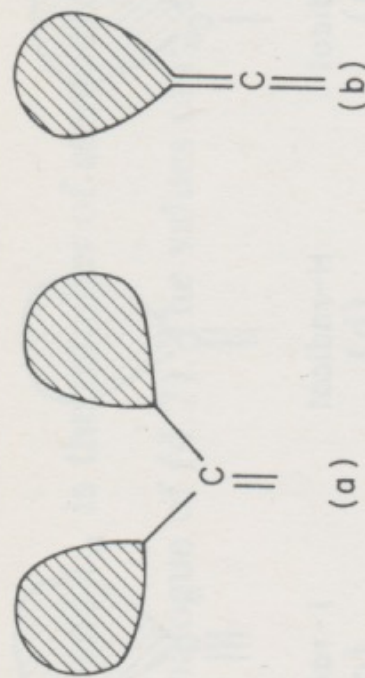


FIG. 7

On closer inspection the H -radicals are seen to be of the two types shown in Fig. 7.

In Fig. 7(a) we have two interchangeable boxes into which G -radicals must be put. This gives us the counting series

$$xyZ(S_2; G(x,y,z))$$

the factor xy being present because of the extra carbon and extra double bond.

In Fig. 7(b) we have a single box into which we must put an H -radical. This gives us $xyH(x,y,z)$. Taking account of the extra double bond and the extra carbon atom introduced we derive the following equation

$$H(x,y,z) = xyZ(S_2; G(x,y,z)) + xyH(x,y,z) \quad (6.5)$$

whence

$$H(x,y,z) = \frac{xy}{1-xy} Z(S_2; G(x,y,z)). \quad (6.6)$$

The I -radicals are of just one type, shown in Fig. 8 and they contribute the counting series $xzG(x,y,z)$. Hence

$$I(x,y,z) = xzG(x,y,z). \quad (6.7)$$

From (6.4), (6.6) and (6.7) we derive

$$G(x,y,z) = 1 + x \left\{ Z(S_3; G(x,y,z)) + \frac{xy}{1-xy} G(x,y,z) \cdot Z(S_2; G(x,y,z)) + xzG(x,y,z) \right\} \quad (6.8)$$

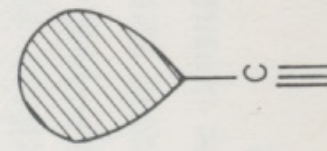


FIG. 8

or

$$G(x,y,z) = 1 + x \left\{ \frac{1}{6} [G^3(x,y,z) + 3G(x,y,z)G(x^2,y^2,z^2) + 2G(x^3,y^3,z^3)] + \frac{1}{2} \frac{xy}{1-xy} [G^3(x,y,z) + G(x,y,z)G(x^2,y^2,z^2)] + xzG(x,y,z) \right\} \quad (6.9)$$

Given $G(x,y,z)$ up to the terms in x^n we can evaluate the right-hand side of (6.9), and hence $G(x,y,z)$, up to the terms in x^{n+1} . Hence $G(x,y,z)$ can be calculated to whatever extent we wish. The calculation is extremely tedious by hand, but has been programmed on a computer, and the results given in Table III were obtained.

TABLE III
Monovalent acyclic radicals (d = double bonds and t = triple bonds)

n	t	0	1	$n=3$	0	1	$n=4$	0	1	2	$n=5$	0	1	2		
0	1	1	0	2	2	0	4	4	1	0	8	10	4			
d	1	1	d	1	3	1	8	3	1	21	14					
				2	1	d	2	5	d	2	20	3				
							3	1		3	7					
										4	1					
n	t	0	1	2	3	$n=7$	0	1	2	3	$n=8$	0	1	2	3	4
0	17	25	12	1	0	39	64	38	7	0	89	166	115	29	1	
1	56	50	7	1	149	166	45	1	398	531	206	13				
2	69	25	2	228	134	7			2	725	587	84				
d	3	37	3	d	3	165	36		d	3	664	261	7			
	4	9		4	60	3			4	326	47					
	5	1		5	11				5	88	3					
				6	1				6	13						
									7	1						
n	t	0	1	2	3	4	$n=10$	0	1	2	3	4	5			
0	211	437	348	114	11	0	507	1157	1040	417	62	1				
1	1068	1656	829	115	1	2876	5076	3103	661	22						
2	2261	2325	577	13	2	6932	8639	3066	222							
3	2505	1470	123		3	9032	7121	1155	13							
d	4	1570	433	7	d	4	6909	2998	162							
	5	570	58		5	3204	648	7								
	6	122	3		6	915	69									
	7	15			7	161	3									
	8	1			8	17										
					9	1										

As a check on this result we see that if we put $y = z = 0$, thus eliminating any radicals with double or triple bonds, we obtain

$$G(x, 0, 0) = 1 + \frac{1}{6}x[G^3(x, 0, 0) + 3G(x, 0, 0)G(x^2, 0, 0) + 2G(x^3, 0, 0)] \quad (6.10)$$

which is simply equation (2.1), thinly disguised.

By putting $y = z = 1$ we obtain a recursion formula for the generating function $T(x) = G(x, 1, 1)$ for all hydrocarbon radicals, irrespective of how many multiple bonds they have. Other special results can be obtained in similar ways, but we must hurry on to consider the problem of enumerating the *unsubstituted* hydrocarbons.

As before, we now consider compounds with a labelled carbon atom. These are of 4 types, as shown in Fig. 9. It is easily verified that the type I compounds are enumerated by

$$xZ(S_4; G(x, y, z)); \quad (6.11)$$

the type II compounds by

$$xZ(S_2; G(x, y, z)) \cdot H(x, y, z); \quad (6.12)$$

the type III compounds by

$$xZ(S_2; H(x, y, z)); \quad (6.13)$$

and the type IV compounds by

$$xG(x, y, z) \cdot I(x, y, z). \quad (6.14)$$

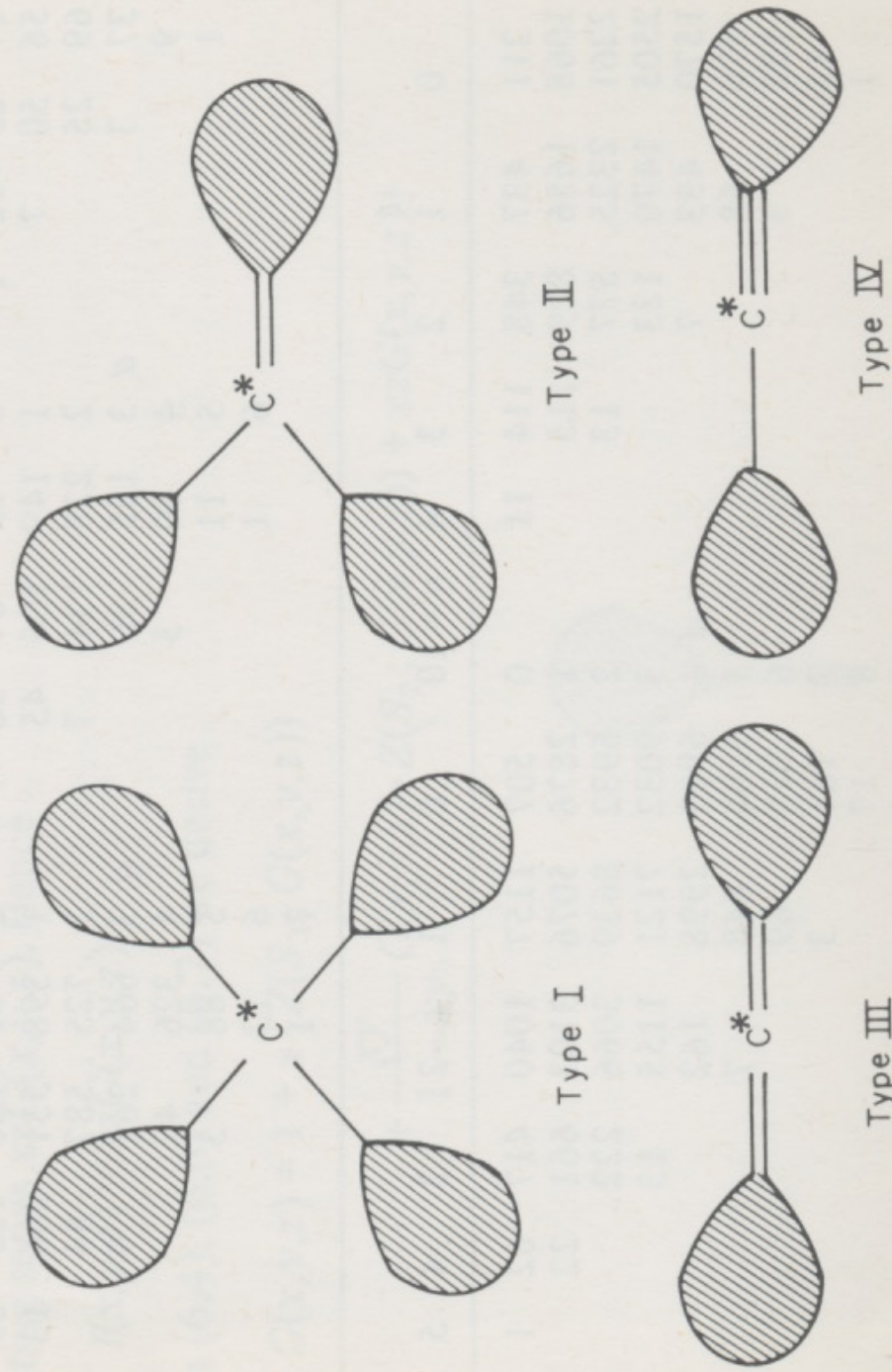


FIG. 9



FIG. 10

All these results follow by application of Pólya's theorem. On adding these four series (6.11-14) we get the counting series for compounds with a labelled carbon atom; we shall denote this series by $K(x, y, z)$.

We now turn to compounds with a labelled carbon-carbon bond. If this bond is single, then our compound is as shown in Fig. 10(a) and we have two interchangeable boxes into which to put a G -radical (other than just hydrogen). The result is

$$Z(S_2; G(x, y, z) - 1). \quad (6.15)$$

However, following our previous method, we shall eventually have to subtract from this the numbers of compounds that are symmetrical about the labelled bond, and now is as good a time as any to do this. These symmetrical compounds are enumerated by $G(x^2, y^2, z^2) - 1$. Subtracting this from (6.15) and, for convenience, writing $G_1(x, y, z)$ in place of $G(x, y, z) - 1$ we obtain

$$\begin{aligned} Z(S_2; G_1(x, y, z)) - G_1(x^2, y^2, z^2) \\ = \frac{1}{2}[G_1^2(x, y, z) - G_1(x^2, y^2, z^2)] \end{aligned} \quad (6.16)$$

This corresponds to the portion $\frac{1}{2}[A_1^2(x) - A_1(x^2)]$ of (4.8).

If the labelled bond is a double bond, as in Fig. 10(b) we proceed similarly but note that in joining two H -radicals in this way we lose a double bond from the total of double bonds that the two radicals had between them; for two double bonds have become one. Hence the analogue of (6.16) is

$$\frac{1}{2y} \{H^2(x, y, z) - H(x^2, y^2, z^2)\} \quad (6.17)$$

where the factor $1/y$ reduces the double bond count by 1 for each term.

Those compounds for which the labelled bond is a triple bond are treated similarly to give

$$\frac{1}{2z} \{I^2(x, y, z) - I(x^2, y^2, z^2)\} \quad (6.18)$$

We now apply the same theorem as before, and we find that the unsubstituted acyclic hydrocarbons are enumerated according to carbon content

and numbers of double and triple bonds by the counting series $L(x,y,z)$, where

$$L(x,y,z) = K(x,y,z) - \frac{1}{2} \{G_1^2(x,y,z) - G_1(x^2, y^2, z^2)\} - \frac{1}{2y} \{H^2(x,y,z) - H(x^2, y^2, x^2)\} - \frac{1}{2z} \{I^2(x,y,z) - I(x^2, y^2, z^2)\} \quad (6.19)$$

and everything on the right-hand side of (6.19) is known.

All the coefficients $L_{n,d,t}$ in $L(x,y,z)$ have been computed for $n \leq 10$ and are given in Table IV.

TABLE IV
Acyclic hydrocarbons (d = double bonds, t = triple bonds)

n	t	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	t	t	t	t	t	
0	1	0	1	0	2	1	0	3	3	2					
d	1	1	1	1	3	1	1	5	4						
		2	1	d	2	2	d	2	6	1					
			3	1	3	1	3	2	4	1					
					4	1	4	1							
$n=6$	0	1	2	3	$n=7$	0	1	2	3	$n=8$	0	1	2	3	4
0	5	7	5	1	0	9	14	11	3	0	18	32	28	10	1
1	13	12	3	1	27	34	12	1	66	95	48	4			
2	16	7		2	44	29	3	2	120	110	22				
d	3	10	1	d	3	32	9	d	3	115	53	3			
4	3				4	15	1	4	62	12					
5	1				5	3		5	21	1					
					6	1		6	4						
					7			7	1						
$n=9$	0	1	2	3	4	$n=10$	0	1	2	3	4	5			
0	35	72	69	28	5	0	75	171	179	88	20	1			
1	153	262	157	29	1	377	718	518	138	8					
2	328	376	120	4	2	901	1245	537	53						
3	367	354	29		3	1196	1074	226	4						
d	4	253	85	3	d	4	964	498	39						
5	100	14			5	491	124	3							
6	28	1			6	160	17								
7	4				7	36	1								
8	1				8	5									
					9	1									

7 ENUMERATION OF UNLABELLED STEREO-HYDROCARBONS

As a final exercise we shall solve the problem for stereo-isomers analogous to that for structural isomers that we have just solved. The reasoning is, for the most part, much the same as before, so that not much need be said about it. There are however a few differences which will be pointed out as occasion arises.

When we examine the monovalent radicals, we see that they are of the same three types as shown in Fig. 6. But in Fig. 6(a) the three boxes may only be permuted cyclically, so that the group of permutations is C_3 , not S_3 . There are not other differences at this stage, and we thus easily obtain the stereo-analogue of (6.4). It is

$$g(x,y,z) = 1 + x\{Z(C_3; g(x,y,z)) + g(x,y,z)h(x,y,z) + i(x,y,z)\} \quad (7.1)$$

where the lower case letters g , h and i correspond, in the case of stereo-isomers, to G , H and I for structural isomers.

The H -radicals are again of the two types shown in Fig. 7, but now the two boxes in Fig. 7(a) cannot be permuted. Thus we have the group E_2 instead of S_2 , and the stereo-analogue of (6.5) becomes

$$h(x,y,z) = xyg^2(x,y,z) + xyh(x,y,z) \quad (7.2)$$

whence

$$h(x,y,z) = \frac{xy}{1-xy} g^2(x,y,z). \quad (7.3)$$

A word of clarification is in order here. We have glibly stated that the double bond of the H -radical in Fig. 7(a) prevents the two boxes from being interchanged (by rotation). But in deriving our final result we shall have situations in which this H -radical will be a part of a larger radical. What if the rest of this larger radical has rotational symmetry? The argument would then break down. Fortunately this cannot happen. On the other side of the carbon to which our H -radical is attached there may be another double bond, and beyond that possibly another, but if we follow this chain of double bonds we must eventually reach a carbon atom at which there are two single bonds. Thus the larger radical, of which the radical of Fig. 6(b) is a part, will appear as in Fig. 11. Now one of the two portions A and B contains the free single bonds for this larger radical; the other does not. Hence there can be no rotational symmetry about the double bond in question.

For the I -radicals we have the same situation as before, and we derive the equation

$$i(x,y,z) = xzg(x,y,z). \quad (7.4)$$

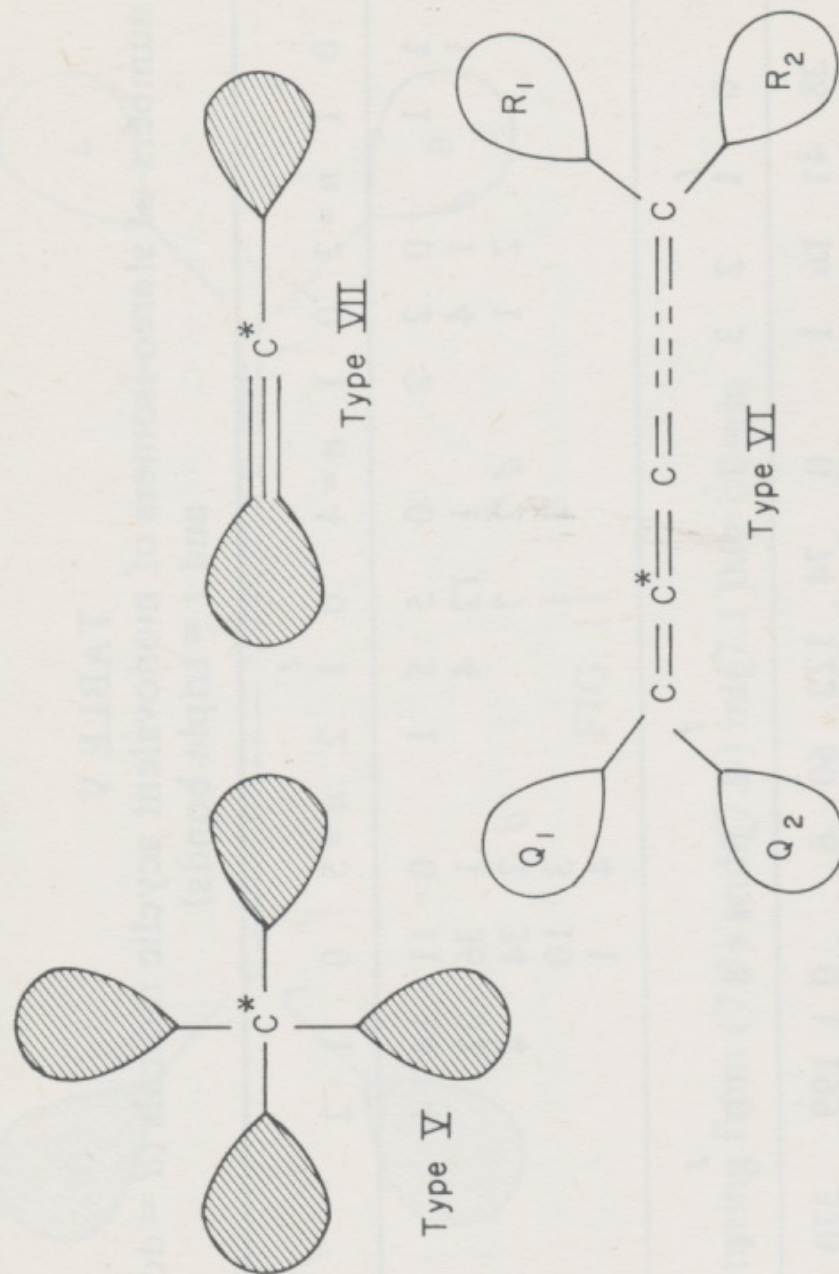


FIG. 12

allow for the addition of $2k$ extra carbon atoms, $2k - 1$ extra double bonds, and the k choices of carbon atom to label. Altogether we have the counting series

$$kx^{2k}y^{2k-1} \cdot \frac{1}{2} \{g^4(x,y,z) + g^2(x^2, y^2, z^2)\}. \quad (7.7)$$

Summing (7.7) over all k we have

$$\sum_{k=1}^{\infty} kx^{2k}y^{2k-1} \cdot \frac{1}{2} \{g^4(x,y,z) + g^2(x^2, y^2, z^2)\} \\ = \frac{x^2y}{2(1-x^2y^2)^2} \{g^4(x,y,z) + g^2(x^2, y^2, z^2)\}. \quad (7.8)$$

If p is odd ($= 2k + 1$) and the atom that is labelled is not the centre one of the chain, then we can take it to be on the left-hand side, and obtain an overall counting series of

$$\sum_{k=1}^{\infty} kx^{2k+1}y^{2k} \cdot \frac{1}{2} \{g^4(x,y,z) + g^2(x^2, y^2, z^2)\} \\ = \frac{x^3y^2}{2(1-x^2y^2)^2} \{g^4(x,y,z) + g^2(x^2, y^2, z^2)\} \quad (7.9)$$

since now $2k + 1$ extra carbons and $2k$ extra double bonds are introduced.

If $p = 2k + 1$ and we label the central carbon atom, then we open up the possibility of symmetry about a 'vertical' axis (i.e. interchange of the left and right ends). The permutations shown in Fig. 13 are then possible.

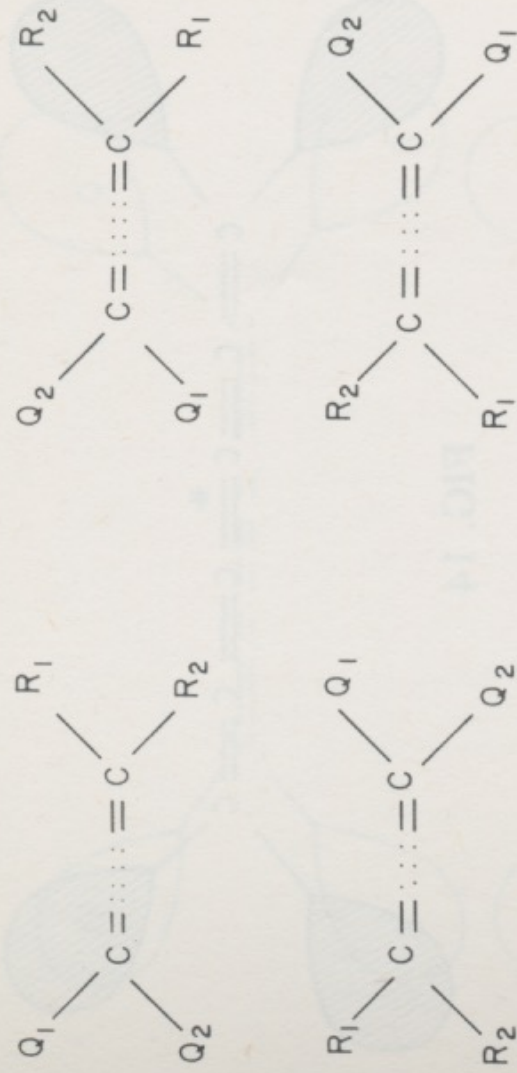


FIG. 13

The cycle index of this group of permutations is $\frac{1}{4}(s_1^4 + 3s_2^2)$. [Compare the problem of ethylene derivatives, and equation (5.3).] Thus in this case we get an additional counting series of

$$\sum_{k=1}^{\infty} x^{2k+1}y^{2k} \cdot \frac{1}{4} \{g^4(x,y,z) + 3g^2(x^2, y^2, z^2)\} \quad (7.10)$$

$$= \frac{x^3y^2}{4(1-x^2y^2)} \{g^4(x,y,z) + 3g^2(x^2, y^2, z^2)\}$$

By adding up the contributions (7.6), (7.8), (7.9) and (7.10) we obtain the counting series for all labelled stereo-isomers, this being the analogue of the series $K(x,y,z)$ for the structural isomer problem. However, since some simplification can be carried out later, we shall not collect together these several series at present.

As before, we now count the stereo-isomers of hydrocarbons with a labelled bond. Those compounds for which the labelled bond is single or triple give no trouble, and if we subtract the numbers of compounds for which the bond in question is symmetric, we obtain

$$\frac{1}{2} [g_1^2(x,y,z) - g_1(x^2, y^2, z^2)] \quad (7.11)$$

where $g_1(x,y,z) = g(x,y,z) - 1$, and

$$\frac{1}{2z} [i^2(x,y,z) - i(x^2, y^2, z^2)]$$

as the analogues of (6.16) and (6.18) respectively. The compounds for which the labelled bond is double present more difficulty, and we consider again the diagram of Fig. 12 (Type VI).

If $p = 2k + 1$, arrange to have the labelled double bond on the left. There are

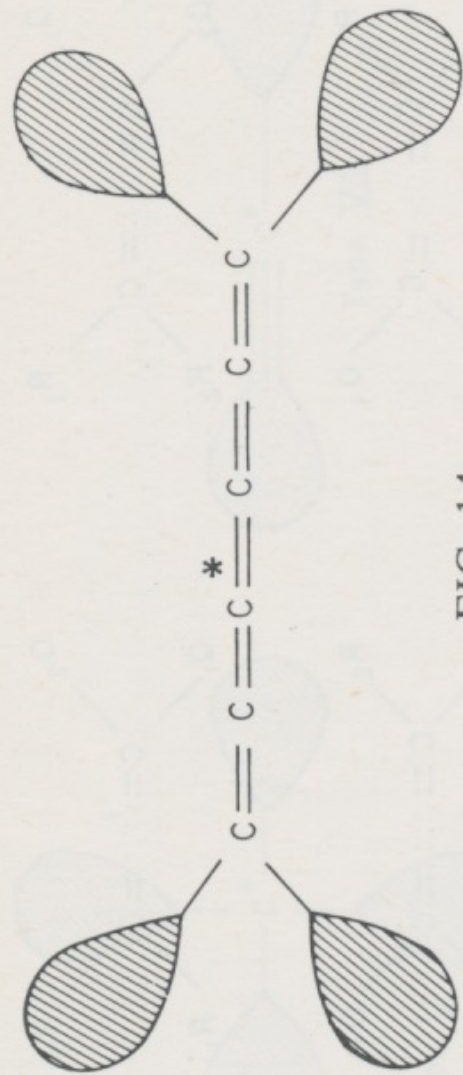


FIG. 14

k ways of choosing it. The group is $S_2 [E_2]$ and as before we obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} kx^{2k+1}y^{2k} \cdot \frac{1}{2} \{g^4(x,y,z) + g^2(x^2,y^2,z^2)\} \\ &= \frac{x^3y^2}{2(1-x^2y^2)^2} \{g^4(x,y,z) + g^2(x^2,y^2,z^2)\}. \end{aligned}$$

Note that this term will cancel with (7.9) when we eventually collect the terms together.

If $p = 2k$ and the labelled bond is not the centre one, arrange for it to be on the left. There are now $k - 1$ choices and we obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} (k-1)x^{2k}y^{2k-1} \cdot \frac{1}{2} \{g^4(x,y,z) + g^2(x^2,y^2,z^2)\} \\ &= \frac{x^4y^3}{2(1-x^2y^2)^2} \{g^4(x,y,z) + g^2(x^2,y^2,z^2)\} \end{aligned} \quad (7.14)$$

If the labelled bond is the centre one (Fig. 14) the requisite group is now that having cycle-index $\frac{1}{4}(s_1^4 + 3s_2^2)$ (compare Fig. 13), and we obtain the series

$$\begin{aligned} & \sum_{k=1}^{\infty} x^{2k}y^{2k-1} \frac{1}{4} \{g^4(x,y,z) + 3g^2(x^2,y^2,z^2)\} \\ &= \frac{x^2y}{4(1-x^2y^2)} \{g^4(x,y,z) + 3g^2(x^2,y^2,z^2)\} \end{aligned} \quad (7.15)$$

We now consider those compounds having a labelled double bond which is symmetric, which can occur only if $p = 2k$ and the centre bond is labelled. Given two different G -radicals Q and R to place at one end of the chain of carbons, we can construct *two* molecules of this type, as shown in Fig. 15.

These are different (*cis* and *trans* isomerism) but each has an automorphism group mapping the central bond end-for-end on itself. However if Q and R are the same there is clearly only one such molecule. Hence the problem is equivalent to that of placing a figure in each of two *distinguishable* boxes, since, for a given choice of figures we have two ways of placing the figures if they

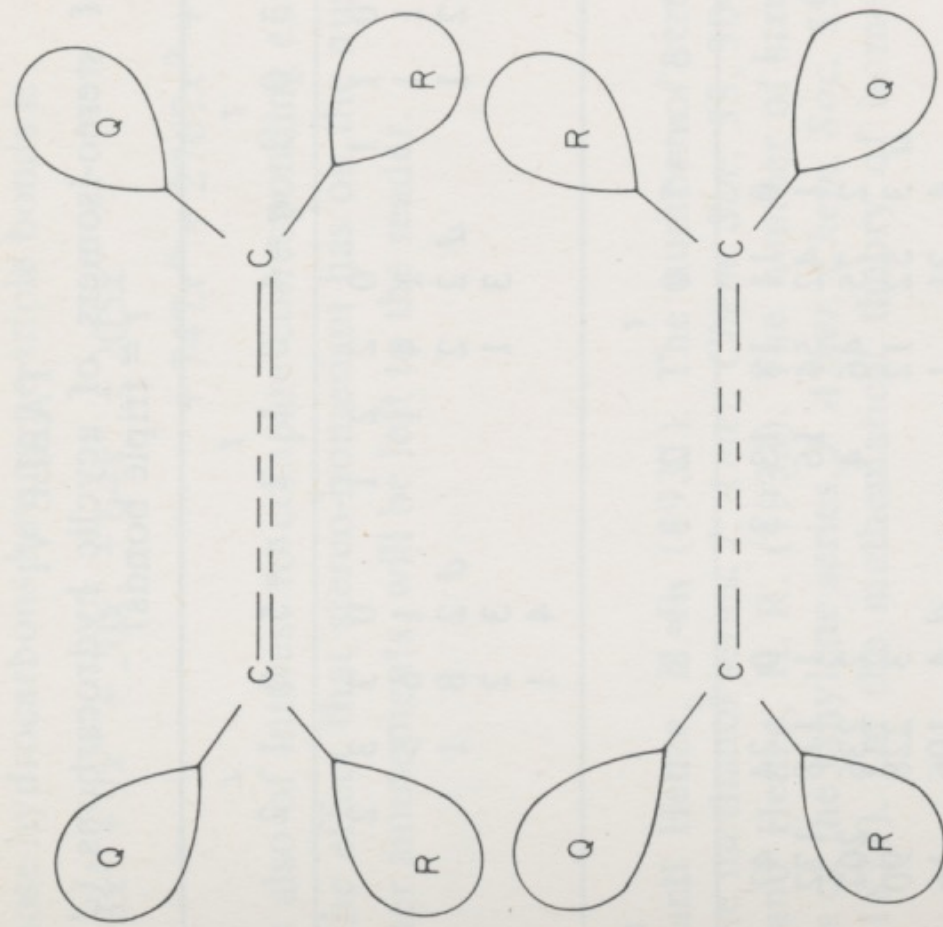


FIG. 15

are different, but only one if they are identical. Hence the number of compounds for which the four G -radicals have parameters n , d and t is the coefficient of $x^n/2y^d/z^t/2$ in $g^2(x,y,z)$. Allowing for the $2k$ extra carbon atoms and the $2k - 1$ double bonds we obtain the series

$$x^{2k}y^{2k-1} \{g^2(x^2,y^2,z^2)\}$$

On summing for $k = 1$ upwards we get

$$\frac{x^2y}{1-x^2y^2} \{g^2(x^2,y^2,z^2)\} \quad (7.16)$$

To obtain the required counting series $l(x,y,z)$ —analogous to $L(x,y,z)$ in (6.19)—we must add (7.6), (7.8), (7.9) and (7.10), subtract (7.11), (7.12), (7.13), (7.14) and (7.15), and finally add (7.16). Two of these terms cancel, as already remarked, and the remainder, after some simplification, yield the following result:

$$\begin{aligned} l(x,y,z) &= xZ(A_4; g(x,y,z)) + \frac{1}{2}x^2z \{g^2(x,y,z) - g(x^2,y^2,z^2)\} \\ &+ \frac{x^2y}{2(1-x^2y^2)} \{g^4(x,y,z) + g^2(x^2,y^2,z^2)\} \\ &- \frac{x^2y}{4(1-xy)} \{g^4(x,y,z) + 3g^2(x^2,y^2,z^2)\} \\ &- \frac{1}{2} \{g_1^2(x,y,z) - g_1(x^2,y^2,z^2)\} \\ &+ \frac{x^2y}{1-x^2y^2} \{g^2(x^2,y^2,z^2)\}. \end{aligned}$$

TABLE VI

The numbers of stereo-isomers of acyclic hydrocarbons (d = double bonds, t = triple bonds)

n	t	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$		
0	1	1	0	2	2	1	0	3	3	2						
d	1	1	1	1	4	1	1	6	5							
		2	2	2	2	2	2	8	1							
			3	1	3	2	3	2								
					4	1	4	1								
$n=6$	t	0	1	2	3	$n=7$	0	1	2	3	$n=8$	0	1	2	3	4
0	5	8	5	1	0	11	18	13	3	0	24	47	37	11	1	
1	18	17	4	1	42	54	16	1	118	172	76	5				
2	23	9	2	75	46	4	2	236	207	32						
d	3	14	1	d	3	52	12	3	228	90	4					
4	3		4	21	1	d	4	106	16							
5	1		5	3	5	30	1									
			6	1	6	4	7	1								
$n=9$	t	0	1	2	3	4	$n=10$	0	1	2	3	4	5			
0	55	123	108	36	5	0	136	338	325	135	23	1				
1	314	544	286	42	1	895	1717	1102	238	11						
2	758	832	217	5	2	2418	3214	1164	86							
3	868	532	44		3	3356	2720	443	5							
d	4	547	149	4	4	5784	1108	60								
5	180	19		d	5	1135	223	4								
6	40	1		6	295	23										
7	4			7	52	1										
8	1			8	87											
				9	11											

This, then, is the counting series for stereo-isomers of the acyclic hydrocarbons. Values of the coefficients in this series are given in Table VI.

From these values we can deduce some of the results that we had earlier. If we pick out the coefficients of the terms independent of y and z we get the series $a(x)$. The terms of the form $x^n y$ will give us the counting series for the ethylene derivatives, i.e. (5.3); those of the form $x^n z$ will give the counting series for the acetylene derivatives, i.e. $xb(x)$.

Other results can be obtained. Thus the counting series for the stereo-isomers of all acyclic hydrocarbons is

$$x + 3x^2 + 4x^3 + 13x^4 + 31x^5 + 109x^6 + 372x^7 + 1446x^8 + 5714x^9 + 27100x^{10} + \dots$$

while that for those hydrocarbons having no triple bonds is

$$x + 2x^2 + 3x^3 + 9x^4 + 20x^5 + 64x^6 + 205x^7 + 747x^8 + 2767x^9 + 14159x^{10} + \dots$$

and so on. It is also of interest to compare corresponding entries in Tables IV and VI, to see the effect that stereo-isomerism has on the numbers of isomers; but this, and other amusements, will be left to the reader.

REFERENCES

- Blair, C. M. and Henze, H. R. (1931). The number of structurally isomeric alcohols of the methanol series. *J. Amer. Chem. Soc.* **53**, 3042-3046.
- Blair, C. M. and Henze, H. R. (1933). The number of structurally isomeric hydrocarbons of the ethylene series. *J. Amer. Chem. Soc.* **55**, 680-686.
- Cayley, A. (1874). On the mathematical theory of isomers. *Phil. Mag.* **47**, 444-446.
- Harary, F. and Norman, R. (1960). Dissimilarity characteristic theorems for graphs. *Proc. Amer. Math. Soc.* **11**, 332-334.
- Perry, D. (1932). The number of structural isomers of certain homologs of methane and methanol. *J. Amer. Chem. Soc.* **54**, 2918-2920.
- Pólya, G. (1938). Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Math.* **68**, 145-254.
- Read, R. C. (1972). Some recent results in chemical enumeration. *In: Graph Theory and Applications* (Eds. Y. Alavi, D. R. Lick, and A. T. White). Springer-Verlag, Berlin.
- Masinter, L. M., Sridhavan, N. S., Lederburg, J. and Smith, D. H. (1974) *J. Amer. Chem. Soc.* **96**, 7702.

A 366 73

A 366 73