Journal of Integer Sequences, Vol. 2 (1999), Article 99.1.1

On Cayley's Enumeration of Alkanes (or 4-Valent Trees)

E. M. Rains and N. J. A. Sloane
Information Sciences Research
AT&T Shannon Lab
Florham Park, NJ 07932-0971

Email addresses: rains@research.att.com, njas@research.att.com

Abstract: Cayley's 1875 enumerations of centered and bicentered alkanes (unlabeled trees of valency at most 4) are corrected and extended — possibly for the first time in 124 years.

1. Introduction

In 1875 Cayley attempted to enumerate alkanes C_nH_{2n+2}, or equivalently n-node unlabeled trees in which each node has degree at most 4, and published a short note [Cay75] containing the table:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	
centered	1	0	1	1	2	2	6	9	20	37	86	183	419	(1)
bicentered	0	1	0	1	1	3	3	9	15	38	73	174	380	(2)
total	1	1	1	2	3	5	9	18	35	75	159	357	799	(3)

(The terms "centered" and "bicentered" are defined below.) This table was reproduced by Busacker and Saaty in 1965 [BuS65], and the three sequences were included in [HIS].

In fact the last two columns are in error, as had already been pointed out by Herrmann in 1880 [Her80]. Herrmann uses a different method from Cayley, and gives the correct values 355 (for $n = 12$) and 802 (for $n = 13$) for sequence (3). However, neither in [Her80] nor in his two later notes [Her97], [Her98] does he mention sequences (1) and (2).

The alkane sequence (3) is also discussed in the works by Schiff [Sch75], Losanitsch [Los97], [Los97a], Henze and Blaire [HeB31], Perry [Per32], Polya [Polya36], [Polya37], Harary and Norman [HaN60], Lederberg [Led69], Read [Rea76], Robinson, Harary and Balaban [RoHB76], and Bergeron, Labelle and Leroux [BeLL98]. The simplest generating function is due to Harary and Norman (see Section 4 of [Rea76] or p. 289 of [BeLL98]). However, none of these authors use Cayley's method, and as far as we can tell none of them discuss sequences (1) and (2).

In 1988 R. K. Guy wrote to N.J.A.S., pointing out that there were errors in these three sequences, and suggested that Polya counting theory be used to extend (1) and (2). (The correct version of (3), sequence A000602, was already present in [HIS].) To do so is the goal of the present note.

We confess to having another, more ignoble reason for wishing to extend sequences (1) and (2). The sequences in the data-base [EIS] are numbered (A000001, A000002, A000003, ...), and several people have suggested that the "diagonal" sequence, whose nth term is the nth term of A_n, should be added to [EIS]. The fact that (1) is sequence A000022 provided additional motivation for extending it to at least the 22nd term! (The "diagonal" sequence is now in the data-base, sequence A031135, as is the even less well-defined A037181 whose nth term is 1 + nth term of A_n.)

The hard part was determining exactly what Cayley was attempting to count, since [Cay75] is somewhat unclear, and contains many typographical errors. Once the problem was identified, it turned out to be quite easy to calculate these sequences — so in fact it is very likely that this has been done in the 124 years since [Cay75] appeared. But we have been unable to find any record of it in the literature.

2. Generating functions

A tree of diameter $2m$ has a unique node called the *center*, at the midpoint of any path of length $2m$. A tree of diameter $2m+1$ has a unique pair of nodes called *bicenters*, at the middle of any path of length $2m+1$. These terms were introduced by Jordan around 1869 ([Har69], p. 35).

Cayley's approach [Cay75] to counting alkanes uses the notions of center and bicenter to reduce the problem to simpler questions about rooted trees. This turns out to be an awkward way to attack the problem (since the notion of diameter is irrelevant), and may explain why no one else has used this approach.

It is simpler to make use of the notion of "centroid" and "bicentroid", also due to Jordan (see Harary [Har69], p. 36, for the definition). In 1881 Cayley [Cay81] found recurrences for the numbers of n-node trees with a centroid (sequence A000676) and with a bicentroid (A000677), which gave him a simpler way to enumerate unrooted trees (A000055). However, as far as we know Cayley did not use the centroid/bicentroid method to enumerate alkanes (A000602). This was apparently first done by Polya [Polya36], [Polya37] in 1936.

However, our concern here is with centered and bi-centered trees.

We will say that a tree is *k-valent* if the degree of every node is at most k. Alkanes are precisely the 4-valent trees.

We wll also consider rooted trees, and define a *b-ary* rooted tree to be either the empty tree or a rooted tree in which the out-degree of every node (the valency excluding the edge connecting it to the root) is at most b. This generalizes the notion of a *binary* rooted tree, the case $b = 2$, which is either the empty tree or a rooted tree in which every node has 0, 1 or 2 sons. (The literature contains several other definitions of binary and b-ary trees. These terms sometimes refer specifically to planar trees. Our trees are not planar, and in particular there is no notion of right or left.)

We will find generating functions for centered and bicentered k-valent trees.

Fix k, and let $T_{h,n}$ be the number of $(k-1)$-ary rooted trees with n nodes and height at most h. (The height of a node in a rooted tree is the number of edges joining the node to the root.) By convention the empty tree has height -1.

Let $T_h(z) = \sum_{n \ge 0} T_{h,n} z^n$. Then $T_{-1}(z) = 1$, $T_0(z) = 1 + z$, and for $h > -1$,

$$T_{h+1}(z) = 1 + z\, \mathbf{S}_{k-1}(T_h(z)), \qquad (4)$$

where $\mathbf{S}_m(f(z))$ denotes the result of substituting $f(z)$ into the cycle index for the symmetric group of order $m!$. For example,

$$\mathbf{S}_3(f(z)) = (f(z)^3 + 3 f(z) f(z^2) + 2 f(z^3)) / 3!.$$

Equation (4) holds because if we remove the root and adjacent edges from a rooted tree of height $h+1$ we are left with an unordered $(k-1)$-tuple of trees of height h.

Let $C_{2h,n}$ be the number of centered k-valent trees with n nodes and diameter $2h$, and let $C_{2h}(z) = \sum_{n \ge 0} C_{2h,n} z^n$.

By deleting the center node and adjacent edges, we see that any such tree corresponds to an unordered k-tuple of $(k-1)$-ary rooted trees of height at most $h-1$, at least two of which have height exactly $h-1$. Therefore

$$C_{2h}(z) = (1 + z\, \mathbf{S}_k(T_{h-1}(z))) - (1 + z\, \mathbf{S}_k(T_{h-2}(z))) - (T_{h-1}(z) - T_{h-2}(z))(T_{h-1}(z) - 1). \qquad (5)$$

The three expressions in (5) account for the k-tuples of rooted trees of height at most $h-1$, k-tuples of rooted trees of height at most $h-2$, and rooted trees with exactly one subtree at the root with height $h-1$, respectively.

Finally, let C_n denote the number of centered k-valent trees with n nodes, and $C(z) = \sum_{n \ge 0} C_n z^n$. Then

$$C(z) = \sum_{h \ge 0} C_{2h}(z).$$

For $k = 4$ we obtain

$$C(z) = z + z^3 + z^4 + 2 z^5 + 2 z^6 + 6 z^7 + 9 z^8 + 20 z^9 + 37 z^{10} + 86 z^{11} + 181 z^{12} + 422 z^{13} + \dots ,$$

which is the corrected version of Cayley's sequence (1), A000022. (See the table below.)

Bicentered trees are easier to handle. Let $B_{2h+1,n}$ be the number of bicentered k-valent trees with n nodes and diameter $2h+1$, let $B_{2h+1}(z) = \sum_{n \ge 0} B_{2h+1,n} z^n$, let B_n be the total number of bicentered k-valent trees with n nodes, and let $B(z) = \sum_{n \ge 0} B_n z^n$. Since a bicentered tree corresponds to an unordered pair of $(k-1)$-ary rooted trees of height exactly h, we have

$$B_{2h+1}(z) = \mathbf{S}_2(T_h(z) - T_{h-1}(z)) ,$$

and then

$$B(z) = \sum_{h \ge 0} B_{2h+1}(z) .$$

For $k = 4$ we obtain

$$B(z) = z^2 + z^4 + z^5 + 3 z^6 + 3 z^7 + 9 z^8 + 15 z^9 + 38 z^{10} + 73 z^{11} + 174 z^{12} + 380 z^{13} + \dots ,$$

Cayley's sequence (2), A000200 (which as it turns out was correct).

The generating function for alkanes (A000602) is then

$$C(z) + B(z) = z + z^2 + z^3 + 2 z^4 + 3 z^5 + 5 z^6 + 9 z^7 + 18 z^8 + 35 z^9 + 75 z^{10} + 159 z^{11} + 355 z^{12} + 802 z^{13} + \dots ,$$

in agreement with Henze and Blair [HeB31] (except that the value they give for $n = 19$, 147284, is incorrect: it should be 148284). Further terms are shown in the following table:

Table: Numbers of centered, bicentered and unrestricted 4-valent trees with *n* nodes

n	centered	bicentered	total
	(A000022)	(A000200)	(A000602)
1	1	0	1
2	0	1	1
3	1	0	1
4	1	1	2
5	2	1	3
6	2	3	5
7	6	3	9
8	9	9	18
9	20	15	35
10	37	38	75
11	86	73	159
12	181	174	355
13	422	380	802
14	943	915	1858
15	2223	2124	4347
16	5225	5134	10359
17	12613	12281	24894
18	30513	30010	60523
19	74883	73401	148284
20	184484	181835	366319
21	458561	452165	910726
22	1145406	1133252	2278658
...	...	...	...

If we set $k = 3$ in the above formulae (corresponding to centered, bicentered and unrestricted 4-valent trees), we obtain sequences A000675, A000673 and A000672, for which the initial terms were (correctly) published by Cayley in another 1875 paper [Cay75a], and further terms were computed by R. W. Robinson in 1975 [Rob75].

For $k = 5$ and 6 the resulting sequences (A036648, A036649, A036650, A036651, A036652, A036653) appear to be new.

References

[BeLL98] F. Bergeron, G. Labelle and P. Leroux, *Combinatorial Species and Tree-Like Structures*, Camb. Univ. Press, 1998, see p. 290.

[BuS65] R. G. Busacker and T. L. Saaty, *Finite Graphs and Networks*, McGraw-Hill, NY, 1965, see p. 201.

[Cay75] A. Cayley, Ueber die analytischen Figuren, welche in der Mathematik Bäume genannt werden und ihre Anwendung auf die Theorie chemischer Verbindungen, *Ber. deutsch. chem. Ges.*, **8** (1875), 1056-1059.

[Cay75a] A. Cayley, On the analytic forms called trees, with applications to the theory of chemical combinations, *Reports British Assoc. Adv. Sci.*, **45** (1875), 257-305 = *Math. Papers*, Vol. 9, pp. 427-460 (see p. 451).

[Cay81] A. Cayley, On the analytical forms called trees, *Amer. J. Math.*, **4** (1881), 266-268.

[Har69] F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969.

[HaN60] F. Harary and R. Z. Norman, Dissimilarity characteristic theorems for graphs, *Proc. Amer. Math. Soc.*, **54** (1960), 332-334.

[HeB31] H. R. Henze and C. M. Blair, The number of isomeric hydrocarbons of the methane series, *J. Amer. Chem. Soc.*, **53** (1931), 3077-3085.

[Her80] F. Hermann, Ueber das Problem, die Anzahl der isomeren Paraffine der Formel $C_nH_{2\,n+2}$ zu bestimmen, *Ber. deutsch. chem. Ges.*, **13** (1880), 792. [Both the author's name and the chemical formula are incorrect.]

[Her97] F. Herrmann, Ueber das Problem, die Anzahl der isomeren Paraffine von der Formel C_nH_{2n+2} zu bestimmen, *Ber. deutsch. chem. Ges.*, **30** (1897), 2423-2426.

[Her98] F. Herrmann, Entgegnung, *Ber. deutsch. chem. Ges.*, **31** (1898), 91.

[Led69] J. Lederberg, Topology of molecules, pp. 37-51 of *The Mathematical Sciences*, M.I.T. Press, Cambridge, MA, 1969.

[Los97] S. M. Losanitsch, Die Isomerie-Arten bei den Homologen der Paraffin-Reihe, *Ber. deutsch. chem. Ges.*, **30** (1897), 1917-1926.

[Los97a] S. M. Losanitsch, Bemerkungen zu der Hermannschen Mittheilung: Die Anzahl der isomeren Paraffine, *Ber. deutsch. chem. Ges.*, **30** (1897), 3059-3060.

[Per32] D. Perry, The number of structural isomers of certain homologs of methane and methanol, *J. Amer. Chem. Soc.*, **54** (1932), 2918-2920.

[Polya36] G. Polya, Algebraische Berechnung der Anzahl der Isomeren einiger organischer Verbindungen, *Zeit. f. Kristall.*, **93** (1936), 415-443.

[Polya37] G. Polya, Kombinatorische Abzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68** (1937), 145-254. Translated as G. Polya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*, Springer-Verlag, NY, 1987.

[Rea76] R. C. Read, The enumeration of acyclic chemical compounds, pp. 25-61 of A. T. Balaban, ed., *Chemical Applications of Graph Theory*, Academic Press, NY, 1976.

[Rob75] R. W. Robinson, personal communication, 1975.

[RoHB76] R. W. Robinson, F. Harary and A. T. Balaban, The numbers of chiral and achiral alkanes and mono substituted alkanes, *Tetrahedron*, **32** (1976), 355-361.

[Sch75] H. Schiff, Zur Statistik chemischer Verbindungen, *Ber. deutsch. chem. Ber.*, **8** (1875), 1542-1547.

[HIS] N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, NY, 1973.

[EIS] N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, published electronically at http://oeis.org.

(Concerned with sequences A000001, A000002, A000003, A000022, A000055., A000200, A000602, A000672, A000673, A000675, A000676, A000677, A031135, A036648, A036649, A036650, A036651, A036652, A036653, A037181.)

Received Aug. 13, 1998; published in Journal of Integer Sequences Jan. 10, 1999.
