ELSEVIER

Enumeration and 3D representation of the stereo-isomers of alkane molecules

Nik Lygeros[a], Paul-Valère Marchand[b], Marc Massot[b,*]

[a]*Institut Girard Desargues – UMR 5028, Université Claude Bernard, Lyon I, 69622 Villeurbanne Cedex, France*
[b]*MAPLY - UMR 5585, Laboratoire de Mathématiques Appliquées de Lyon, Université Claude Bernard, Lyon 1, 69622 Villeurbanne Cedex, France*

Abstract

The theory of chemical enumeration can be considered to have reached maturity with the works by Balaban, and more specifically for the enumeration of alkane isomers, with the work of Pólya and Read, the key points of which rely on the theory of generating function in relation with symmetry properties. The present paper, after briefly recalling the history and the basis of these theories, aims at indicating a new direction of research, which covers a new field of applications beyond the classical results of chemical enumeration, making use of the recent developments of computer algebra. This new direction aims at providing the fundamentals in order to generate explicitly 3D configurations of the stereo-isomers of alkane molecules and consider only some subclasses of the whole set of alkanes which are encountered in most applications. This capability associated with the obtained degree of symmetry of the generated stereo-isomers is of primary importance for applications in thermochemistry and chemical kinetics.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Chemical enumeration; Computer algebra; Stereo-isomers; Alkane molecules

1. Introduction

1.1. From graphs to molecules

A graph is the application of a set on itself. However, the geometrical realization of a graph is more appealing, namely a collection of points and of lines joining some of

---

* Corresponding author. Tel.: +33 4 72 43 10 08; fax: +33 4 72 4 80 53.
  E-mail address: massot@maply.univ-lyon1.fr (M. Massot).

these points either to other points or to themselves. The name graph originates from this geometrical realization and was introduced by Sylvester. Graphs are topological rather than geometrical objects, having as the most important feature the vicinity relationships between points. The correspondence between graphs and chemical categories has found numerous applications in chemistry: a graph corresponds to a molecule, i.e., points symbolize atoms and lines symbolize chemical bonds. Graphs gave Cayley the incentive to develop a procedure for counting the constitutional isomers of alkanes (Pólya, 1937), that is isomers without regard of the three-dimensional structure of the molecules. Later it led Pólya toward the discovery of his powerful counting theorem, which can be applied even to stereo-chemical problems, where the three-dimensional configurations are taken into account. Thus graph theory, the beginning and development of which takes its origin in chemistry, is well suited for solving chemical problems, both by the high degree of abstraction of such concepts as points, lines, neighbors (Harary and Palmer, 1973), as well as by the combinatorial derivation of graph-theoretical concepts which mimic the structure of chemistry viewed as the study of combinations between atoms.

1.2. The basis of chemical enumeration

Graph theory was independently discovered on several occasions and three names deserve special mention—Euler, Kirchhoff, and Cayley. Euler published the first known paper on graph theory in 1736, in which he resolved the Königsberg bridge problem (Biggs et al., 1986). Kirchhoff discovered graphs while solving problems involving the calculation of currents in electrical networks. Organic chemistry became the third breeding ground for graph theory. The best known early organic chemists, who founded the structure theory, were Couper, Butlerov, and Kekulé. They found it convenient to represent a covalent bound between two atoms as a line joining two points; thus every structural formula is a graph. Chemists were then able to predict the number of isomers of alkanes and alkanols of low molecular weights on the basis of simple graphical constructions. The enumeration of the chemical isomers, in particular the constitutional isomers of alkanes C_nH_{2n+2}, was a challenging mathematical problem to which, in 1874–5, Cayley applied the graph concept of a tree put forward by him in 1857.

If only carbon atoms of alkanes are depicted, there is a one-to-one correspondence between these isomers and trees whose points have at most degree 4. To solve the problem of finding the number of such trees, Cayley altered the problem, enumerating first the rooted trees (i.e., an alkane molecule with a labeled carbon atom) and then the unrooted trees. He developed a solution for the initial problem which enabled him to count constitutional isomers of alkanes. However, his results were only correct up to 11 carbon atoms.

Between 1932 and 1934, Blair and Henze succeeded in enumerating all the constitutional isomers of alkanes and alkyls, considering not only constitutional isomers, as Cayley had done, but also stereo-isomers (Jaffe and Orchin, 1965). The methods developed by these authors also yielded the isomer numbers for unsaturated hydrocarbons as well as constitutional isomers of the main types of aliphatic compounds. They realized that it was too optimistic to hope for a formula giving the numbers of isomers in a series as a function of the number of carbon atoms, and so they concentrated instead on setting up recursion

formulae. As we will see in the present paper, such an approach finds its justification in recent years with the development of computer power and the associated computer algebra tools.

1.3. Pólya's contribution

Pólya's Hauptsatz can be viewed as a way of counting functions from a domain into a range of equivalence determined by a particular permutation group acting on the domain. It enables one to express the generating functions for a class of chemical compounds in terms of an appropriate permutation group and another generating function called the figure-counting series. Its generality and ease of application make this result a most powerful tool in enumerative analysis.

Let A be a permutation group over a finite object set X, and for each permutation α in A and each non-negative integer r, let $j_r(\alpha)$ be the number of cycles of r objects in the disjoint cycle decomposition of α. The cycle index of A is the polynomial in the variable $s_1, s_2, s_3, \ldots$ given by the following expression:

$$Z(A) = \frac{1}{|A|} \sum_{\alpha \in A} \prod_r s_r^{j_r(\alpha)}.$$

Pólya's Hauptsatz enumerates classes of functions from X, the object set of the permutation group A, into another set Y of elements called figures. Two such functions f_1 and f_2 are in the same class or are A-equivalent if there is a permutation α in A such that $f_1(\alpha k) = f_2(k)$ for all objects k in X. Often the figures in the set Y have integral weights assigned to them and are also enumerated by a generating function called the figure-counting series, denoted by $c(x) = c_0 + c_1 x + c_2 x^2 + \cdots$, which has as coefficient of x^m the number of elements in Y of weight m. In this case, we can define the weight of a function to be the sum of the weights of the images of all domain elements. In general then, let w be any function from Y into the set of non-negative integers. Any function f from X into Y has as its weight $\sum_{k \in X} w(f(k))$.

Pólya's Hauptsatz (Pólya, 1937; Pólya and Read, 1987). The generating function $C(x)$ which enumerates equivalence classes of functions determined by the permutation group A is obtained by substituting the figure-counting series $c(x)$ in the cycle index $Z(A)$ as follows. Each variable s_r in $Z(A)$ is replaced by $c(s^r)$. Symbolically we write: $C(x) = Z(A; c(x))$.

1.4. Read's contribution

The originality of Read was to combine in Read (1976) the power of Pólya's theory with a remarkable result of Harary and Norman (1960) in order to enumerate the alkane constitutional isomers and stereo-isomers.

The first preliminary result used by Read is the enumeration of alkanes in which one carbon atom has been labeled. Since, in a first step, Read only considers constitutional isomers, the radicals attached to the labeled carbon atom can be permuted by any permutation of S_4. Pólya's theorem is then applied with $A(x)$ as the figure-counting series, and S_4 as the group. Since the cycle index of the group is $\frac{1}{24}(s_1^4 + 6s_1^2 s_2 + 3s_2^2 + 8s_1 s_3 + 6s_4)$
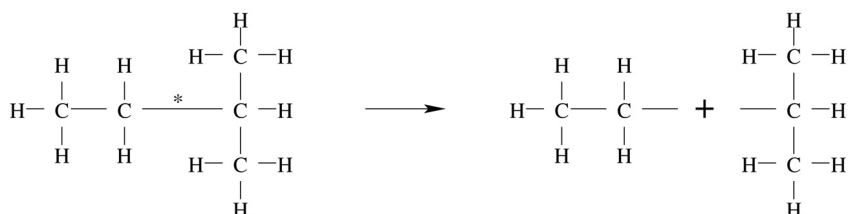
Fig. 1. 'Bond-labeled' molecules and radicals.

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

$$P(x) = \sum_{n=1}^{\infty} P_n x^n = x Z(S_4; A(x))$$

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

where P_n is the number of these compounds with n carbon atoms as a whole.

The second preliminary result required is the enumeration of alkanes in which one carbon-to-carbon valence bond has been labeled, that is, distinguished from the other bonds. This labeling does not correspond in any direct way to a feasible chemical procedure, but is only a means to an end. As with the labeling of a carbon atom, this labeling of a bond gives us a point to start from. If we 'break' the labeled bond in two, we get two alkyles radicals as shown in Fig. 1 (in which the labeled bond is indicated by the asterisk).

Conversely, by joining two alkyl radicals, in any 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-labeled' molecules is a Pólya type problem. Pólya's theorem then gives the configuration counting series

$$Q(x) = \sum_{n=1}^{\infty} Q_n x^n = Z(S_2; A(x) - 1)$$

$$= \frac{1}{2}\{(A(x) - 1)^2 + A(x^2) - 1\}.$$

Let T be any tree. There will be certain one-to-one mappings of the vertex set of T onto itself which leave the tree invariant, i.e., adjacent vertices remain adjacent, etc. These automorphisms form a group. 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 $p^\star$ denote the number of equivalence classes. 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 $q^\star$ be the number of equivalence classes. The result

used by Read, proved by Harary and Norman, is that $p^\star - q^\star + s = 1$ (Harary and Norman, 1960).

Here, $s = 1$ if the tree has a symmetric edge, 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. Let us sum the previous equation over all the alkanes having a given number n of carbon atoms. We have $\Sigma p^\star - \Sigma q^\star + \Sigma s = \Sigma 1$.

By what has just been said, $\Sigma p^\star$ will be the total number of alkanes with a labeled carbon atom; $\Sigma q^\star$ will be the number with a labeled bond; $\Sigma 1$ will be the number of alkanes—the number we are looking for. Thus $\Sigma p^\star$ is the number P_n given by the $P(x)$ formula and $\Sigma q^\star$ is the number Q_n given by the $Q(x)$ formula. If we can determine the number Σs of alkanes with a symmetric bond, we shall know everything, and thus the number we wish to find.

A molecule with a symmetric bond must have an even number of carbon atoms, and as already remarked, the 'half-trees' at the end of the symmetric bond must be the same. Hence, by splitting the symmetric bond Read gets two alkyl radicals for one half and has determined the whole molecule. Hence the number of alkanes having a symmetric edge is $A_{n/2}$. Denoting the required number of alkanes by C_n, we have $P_n - Q_n + A_{n/2} = C_n$.

If we now multiply it by x^n, sum from $n = 1$ upward, and rearrange, we have

$$C(x) = \sum_{n=1}^{\infty} C_n x^n = P(x) - Q(x) + A(x^2) - 1.$$

Since $P(x)$, $Q(x)$ and $A(x)$ are known we can calculate the series $C(x)$ and thus determine the number of alkanes. If we write $A_1(x)$ for $A(x) - 1$, we obtain $C(x) = P(x) - \frac{1}{2}\{A_1^2(x) - A_1(x^2)\}$.

To enumerate the alkanes, having a regard to stereo-isomers, Read first uses alkanes with a labeled carbon atom. It is readily verified that we can permute by any element 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

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

The counting series for alkanes in which a carbon–carbon bond has been labeled is $q(x) = Z(S_2; a(x) - 1) = \frac{1}{2}\left\{[a(x) - 1]^2 + a(x^2) - 1\right\}$.

Read finally uses the theorem to the effect that $p^\star - q^\star + s = 1$, and obtains the result $c(x) = p(x) - q(x) + a(x^2) - 1$, where

$$c(x) = \sum_{n=1}^{\infty} c_n x^n$$

and c_n is the number of stereo-isomers with the formula $C_n H_{2n+2}$.

1.5. Isomers enumeration: from counting series to exhaustive generation

The theory of chemical enumeration can be considered to have reached maturity with the works by Balaban (1976, 1987, 1993), and more specifically for the enumeration of

alkane stereo-isomers, with the work of Read (1972, 1976), the key points of which have been recalled in the previous subsections. The present paper aims at indicating a new direction of research, making use of the recent developments of computer algebra (McKay, 1981; Chaunier and Lygerōs, 1992, 1994; Lygerōs and Mizony, 1996), which covers a new field of applications beyond the classical results of chemical enumeration. This new direction aims at providing the basis in order to generate explicitly 3D configurations of the stereo-isomers of alkane molecules and consider only some subclasses of the whole set of alkanes.

Let us first explain what is meant by a subclass of the whole set of alkanes and why it can be interesting to consider such a subset. In many practical applications ranging from chemical engineering, thermodynamics, up to combustion problems such as Massot et al. (1998); Laurent and Massot (2001), the hydrocarbons under consideration only contain a few carbon atoms. More specifically, the branching complexity of the considered alkanes usually remains very limited. As a consequence, we restrict ourselves, in this paper, to alkanes being constituted of a main chain, on which only methyl (CH_3, i.e., a chain of one carbon atom) and ethyl (C_2H_5, i.e., a chain of two carbon atoms) radicals can be connected; this condition was introduced by Baltanas and will be denoted in the following Baltanas' condition. In the present study, we perform the calculations up to 23 carbons. After having noticed that this subset will rapidly converge to zero density in the set of all alkanes and characterized the represented percentage as a function of the number of carbon atoms, we still remark that this subset is going to represent the overwhelming majority of molecules considered in practical problems. It is then interesting from a mathematical point of view to perform the enumeration of stereo-isomers for this particular subset of alkane molecules, for which the classical tools of counting series will not apply.

Second, the principle of chemical enumeration presented in Balaban (1976) is based on counting series, and thus only delivers stereo-isomer numbers, whereas the geometrical background of the molecules has been included at some point in terms of a permutation subgroup (Pólya and Read, 1987). The purpose of the present study is to perform a chemical enumeration study yielding not only the stereo-isomer numbers, in a global perspective, but also the various three-dimensional configurations with their associated degree of symmetry. This information will be essential in order to evaluate chemical or physical mixture properties based on the spatial structure of the various configurations of the molecule.

The paper is organized as follows. We introduce, in Section 2, an octahedral network paving the three-dimensional space and present the coding dedicated to the subclass of the alkane molecules we have introduced and the associated weight. Section 3 is devoted to the exhaustive enumeration itself and some of the corresponding results are presented. We come to a conclusion in Section 4.

2. Strategy and coding

2.1. Octahedral network in 3D space

The approach used by Read revealed itself as a breakthrough in the field; however, it is implicit and holistic in the sense that it only yields one global numerical piece of information. Our approach is explicit and granular; it is thus fundamentally different from
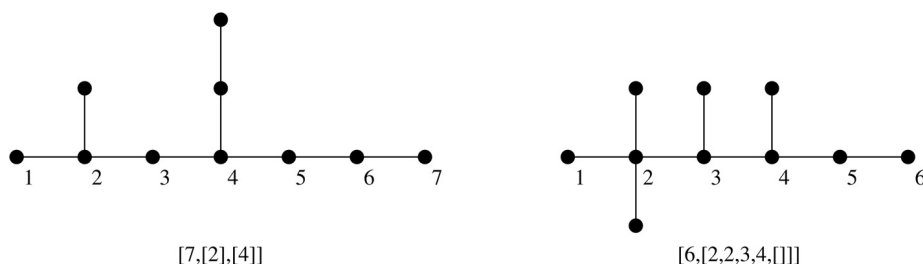
Fig. 2. Two non-isomorphic configurations of $C_{10}H_{22}$.

the preceding one and is rendered possible by the recent developments of the power of computer algebra.

The base of our strategy is to embed the alkane molecules in a centered network of non-regular octahedra paving the three-dimensional Euclidian space. This network allows an easy coding of the molecule as well as an effective isomorphism test through rotations associated with the network. More explicitly, we first of all generate constitutional isomers and classify them in families according to their topological properties; each family is then partitioned into subfamilies based on the associated automorphism group. Finally, we rigidify the obtained structures, place them in the centered network of non-regular octahedra and generate the whole set of stereo-isomers thanks to the exhaustive study of chirality.

2.2. Fundamentals of the coding—constitutional isomers

The selected coding consists of encased lists. This type of object allows a fast access to the data with Maple 7.00, Waterloo Maple Inc. The coding will have only three operands on the first level:

t, size of the principal chain (the number of vertices);
$[m_1, m_2 \ldots, m_p]$, the list of the ranks of the connected methyls;
$[e_1, e_2 \ldots, e_q]$, the list of the ranks of the connected ethyls.

Fig. 2 shows two examples of coding. The flexibility of this code and its effectiveness are due to the fact that only two types of connections are considered. An extension to other to a richer set of alkyl radicals is possible but coding becomes increasingly heavy for a diversity of the "lateral branches".

2.3. The weight of a coding—constitutional isomers

A significant question, leading to an orientation in the programming, was to know how to avoid generating too great a number of molecules. Indeed, as we want to enumerate non-isomorphic configurations, the two graphs of Fig. 3, for example, will have to be regarded as two representations of the same molecule. A way of avoiding the generation of these two configurations is to introduce a *weight* function. The weight of a coding will be defined by the function
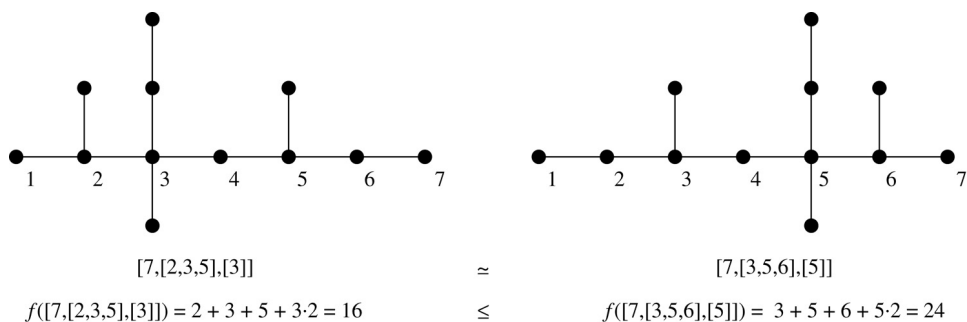
$[7,[2,3,5],[3]]$ $\simeq$ $[7,[3,5,6],[5]]$

$f([7,[2,3,5],[3]]) = 2 + 3 + 5 + 3{\cdot}2 = 16$ $\leq$ $f([7,[3,5,6],[5]]) = 3 + 5 + 6 + 5{\cdot}2 = 24$

Fig. 3. Two isomorphic configurations with different weights.

$$f : [t, [m_1, m_2, \ldots, m_p], [e_1, e_2, \ldots, e_q]] \to \sum_{i=1}^{p} m_i + 2 \sum_{j=1}^{q} e_j.$$

This function calculates a certain barycenter on the connections (a weight of 1 is allocated to methyls and a weight of 2 to ethyls). The program begins and "sets" the connections on the left of the chain; those are then moved gradually and the process stops when the "barycenter" exceeds half of the chain. In this way, it avoids counting twice identical molecules according to whether one reads them starting from the right-hand side or from the left-hand side.

The use of 'topological indices'-like tools will not be pushed further for three reasons. First, for our subclass, we have a very efficient way of generating the set of constitutional isomers, second, the use of the Balaban *J* index or Kier and Hall's total topological state is limited in terms of the number of carbon atoms for our purposes (Randić et al., 2000), and finally this kind of approach has not yet been extended to chiral molecules.

2.4. The 3D coding—stereo-isomers

The centered network of non-regular octahedra is naturally interpreted as Voronoi cells of the carbon atoms constituting two entangled, similar and shifted alkane molecules occupying the whole space. Subsequently, we consider the equivalence class up to discrete carbon–carbon rotations; this equivalence class is supported by the network. A natural representative of this class, because of the Baltanas condition, is obtained by choosing a molecule, the longest chain of which lies in a plane, and by locating the position of the possible branches by "U" for up and "D" for down with respect to this plane. Thus the spatial configuration of the molecule is coded in a canonical way (and Fig. 4 shows two examples of the coding):

t, size of the principal chain (number of vertices);
$[m_1, m_2 \ldots, m_p]$, list of ranks of the connected methyl radicals;
$[\epsilon_1^m, \epsilon_2^m, \ldots, \epsilon_p^m]$, list of position of methyl radicals, where ϵ_i^m belongs to the set $\{U, D\}$. It is chosen according to whether the i-th methyl branch is above the plane of the principal chain or below.
$[e_1, e_2 \ldots, e_q]$, list of ranks of the connected ethyl radicals.
$[\epsilon_1^e, \epsilon_2^e, \ldots, \epsilon_q^e]$, list of position of ethyl radicals, where $\epsilon_i^e \in \{U, D\}$.
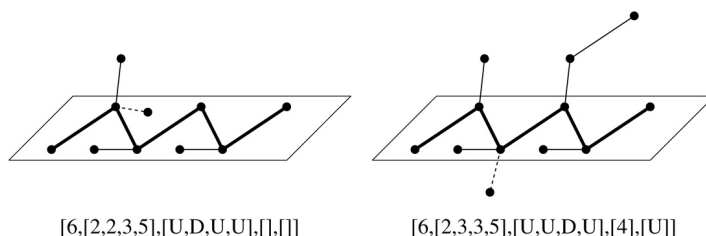
[6,[2,2,3,5],[U,D,U,U],[],[]]   [6,[2,3,3,5],[U,U,D,U],[4],[U]]
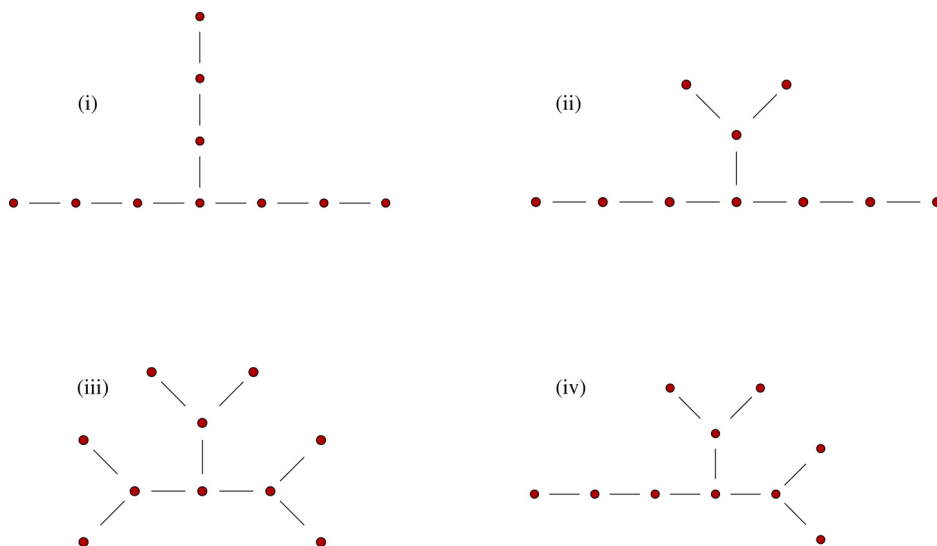
Fig. 4. Examples of 3D codings.

Fig. 5. Molecules with 10 vertices which do not satisfy Baltanas' condition.

3. Enumeration results

3.1. Constitutional isomers

In the tables below, we compare the results obtained by Read, making use of generating functions, with the effective computer algebra method proposed in the present paper under Baltanas' condition. The details of the evaluation of such numbers will be presented below; we first give the final result for comparison purposes.

Baltanas' condition is not restrictive for $n \leq 9$, in such a way that Read's isomer numbers provide an independent certificate in this range of elements. The first graphs with 10 and 11 elements, which we do not enumerate because of Baltanas' condition, are represented in Figs. 5 and 7.

The trees satisfying Baltanas' condition asymptotically represent only a family of zero density as the whole of the quaternary trees which represent exactly the alkane molecules. Table 1 provides the behavior of the number of alkanes satisfying Baltanas' condition with

Table 1
Number of constitutional isomers and comparison

n	Generating functions	Computer algebra under Baltanas' condition	Percentile (%)
1	1	1	100.00
2	1	1	100.00
3	1	1	100.00
4	2	2	100.00
5	3	3	100.00
6	5	5	100.00
7	9	9	100.00
8	18	18	100.00
9	35	35	100.00
10	75	71	94.67
11	159	143	89.94
12	355	295	83.10
13	802	606	75.56
14	1858	1261	67.87
15	4347	2616	60.18
16	10359	5461	52.72
17	24894	11385	45.73
18	60523	23806	39.33
19	148284	49748	33.55
20	366319	104109	28.42
21	910726	217808	23.92
22	2278658	455993	20.01
23	5731580	954512	16.65

respect to the alkane number as a function of the number of carbon atoms n. It is shown to decrease exponentially at infinity and very rapidly even for low values of n as shown in Fig. 6 for both constitutional isomers and stereo-isomers. However, the alkane molecules satisfying Baltanas' condition represent the great majority of the alkane molecules one usually encounters in most chemical applications.

In order to enumerate the considered subclass of alkane molecules, we introduce a granular approach; it consists in a partition of the graphs according to the number and of the type of branches.

The following table provides a typical example of a class in the partition we have performed among the considered alkane molecules in terms of their automorphism group. It highlights the recursive formulae that can be deduced from the various tables. In the following, p represents the number of methyl radicals and q the number of ethyl ones, in agreement with the notations used for coding. In Fig. 8, we give a few examples of molecules with $n = 11$, 12 and 13 having as automorphism groups $S_2 \times S_3$, $(S_2)^4$ and $(S_3)^2$.

Starting from Table 2, one can easily build up recursive formulae for a given topology (for example $p = 2$ and $q = 2$ such as in Table 2) and automorphism group (a given column in Table 2). One can wonder whether or not these formulae make it possible to calculate the number of isomers for alkanes. Fig. 9 shows that new families of graphs appear as the number of vertices increases. This phenomenon suggests the difficulty of tackling the enumeration with a global approach.
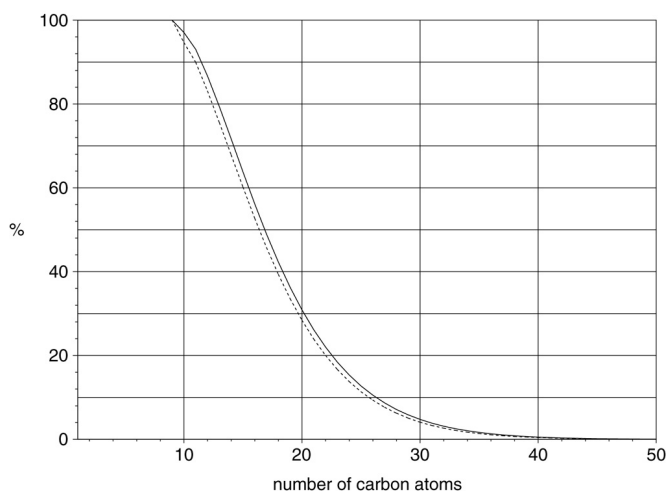
Fig. 6. Percentage of alkane molecules under Baltanas' condition (dashed line: constitutional isomers, solid line: stereo-isomers).

Table 2
Distribution of isomers according to the automorphism group. The lateral branches are composed of 2 methyls and 2 ethyls

$p = 2$ $q = 2$	Automorphism group								
n	S_1	S_2	$(S_2)^2$	S_3	$(S_2)^3$	$S_2 \times S_3$	$(S_2)^4$	$(S_3)^2$	Sum
11	0	0	0	0	0	0	1	1	2
12	0	0	3	0	3	4	0	1	11
13	0	6	15	2	3	8	2	1	37
14	2	30	32	6	7	12	0	1	90
15	14	74	62	12	7	16	2	1	188
16	42	158	93	20	11	20	0	1	345
17	102	276	141	30	11	24	2	1	587
18	204	456	186	42	15	28	0	1	932
19	372	684	252	56	15	32	2	1	1414
20	620	996	311	72	19	36	0	1	2055
21	980	1370	395	90	19	40	2	1	2897
22	1470	1850	468	110	23	44	0	1	3966
23	2130	2406	570	132	23	48	2	1	5312
24	2982	3090	657	156	27	52	0	1	6965
25	4074	3864	777	182	27	56	2	1	8983

3.2. Stereo-isomers

Once the constitutional isomers have been enumerated and characterized, the obtained structures are now embedded in the octahedral network. Only the positions of the side branches are studied, as explained before, and the results are presented in Table 3, and compared with Read's stereo-isomer numbers.
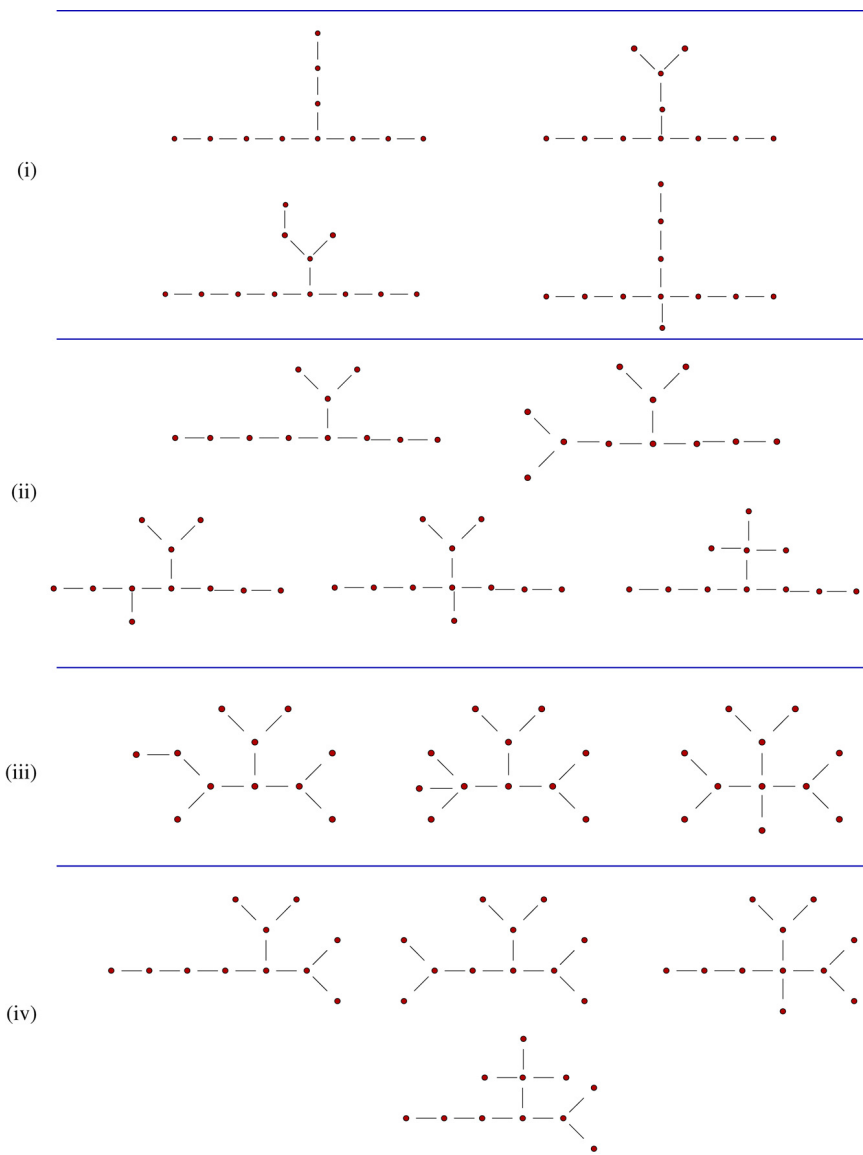
Fig. 7. Molecules with 11 vertices which do not satisfy Baltanas' condition.

As already mentioned for the evaluation of the constitutional isomer numbers, the partition by automorphism group provides simple recurrence formulae for the enumeration of stereo-isomers. The following table is related to the same family as in the example of the previous subsection.

In Fig. 10, as an example, we have represented the 6 stereo-isomers corresponding to $n = 12$ that have $S_2 \times S_3$ as automorphism group. For the sake of legibility, we have only
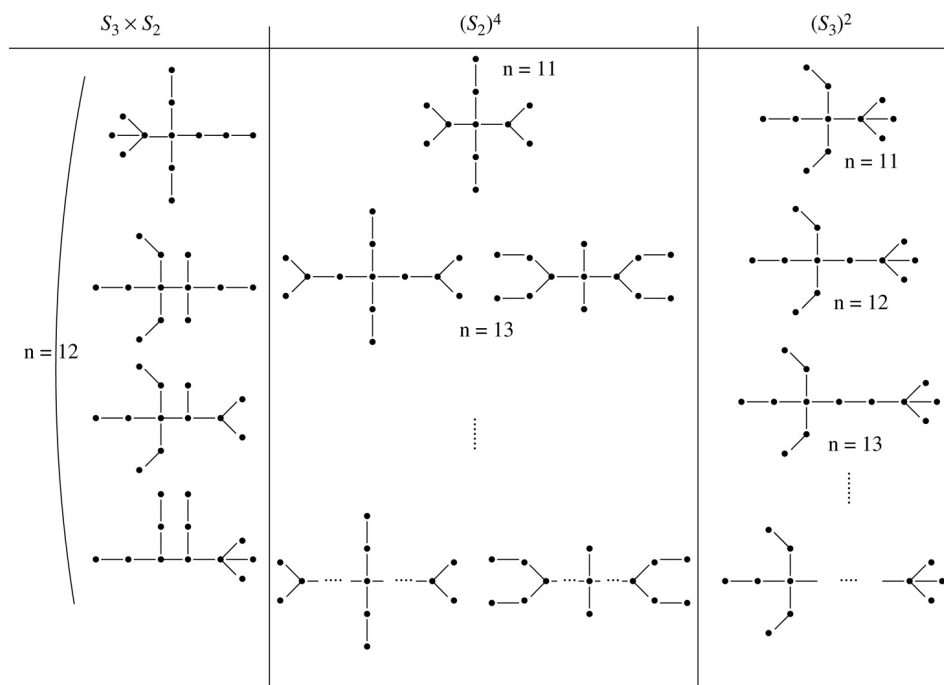
Fig. 8. Examples of molecules having $S_2 \times S_3$, $(S_2)^4$ and $(S_3)^2$ as automorphism groups.

presented a few cases $p = 2$ and $q = 2$ for stereo-isomers in Table 4, and ($p = 4, q = 0$) for the constitutional isomers in Table 6.

As far as the computational cost is concerned, the present result have been obtained on a single Pentium III, 1 GHz computer and it took a little more than 8 h in order to obtain the number of constitutional and stereo-isomers for $n = 23$. The computational cost is presented in detail in Table 5. However, our approach can be first extended to a C program and used on a parallel cluster of more recent PC's or even distributed on various computers. It can then go far beyond $n = 23$, a value we have chosen for illustration and comparison purposes.

4. Conclusion

Our effective approach to the problem of the isomorphic enumeration of stereo-isomers is basically different from Polya's method and Read's method because it is based on an exhaustive generation of the geometric structure as well as optical activity and degree of symmetry of the alkanes and not simply the global number of stereo-isomers.

In our approach, the geometrical structure, which can be interpreted for an alkane molecule as an embedding into a centered network of non-regular octahedra paving the 3D space, does not only result in the definition of formal series and finally in a way of calculating the global number of stereo-isomers as in Read, but all the non-isomorphic stereo-isomers are generated with their degree of symmetry.

Table 3
Number of stereo-isomers and comparison

n	Generating functions	Computer algebra under Baltanas' condition	Percentile (%)
1	1	1	100.0
2	1	1	100.0
3	1	1	100.0
4	2	2	100.0
5	3	3	100.0
6	5	5	100.0
7	11	11	100.0
8	24	24	100.0
9	55	55	100.0
10	136	132	97.1
11	345	321	93.0
12	900	782	86.9
13	2412	1916	79.4
14	6563	4705	71.7
15	18127	11567	63.8
16	50699	28468	56.2
17	143255	70108	48.9
18	408429	172721	42.3
19	1173770	425631	36.3
20	3396844	1049038	30.9
21	9892302	2585793	26.1
22	28972080	6374186	22.0
23	85289390	15713531	18.4

Table 4
Distribution of stereo-isomers according to the automorphism group. The lateral branches are composed of 2 methyls and 2 ethyls

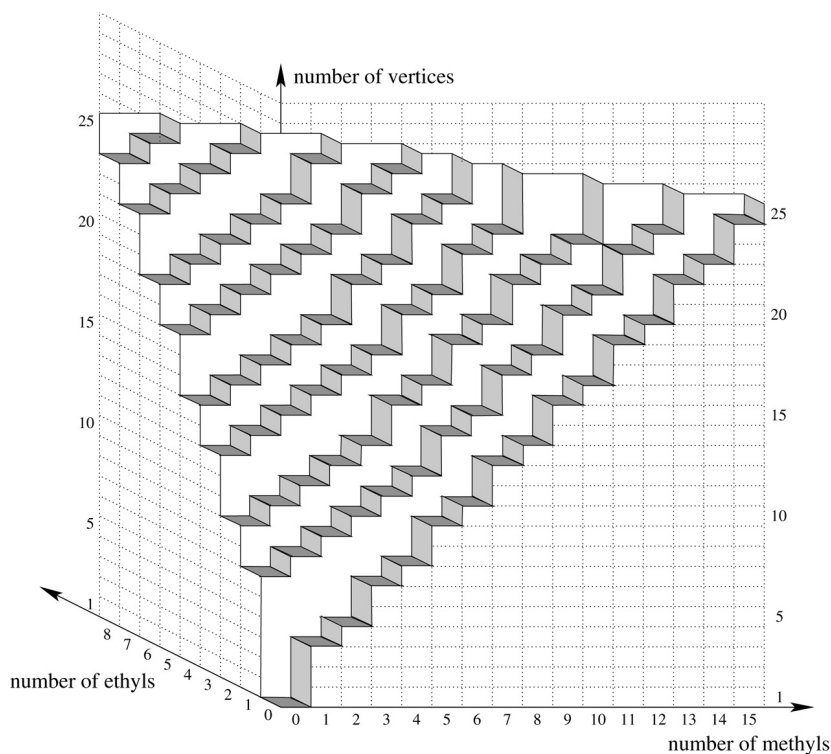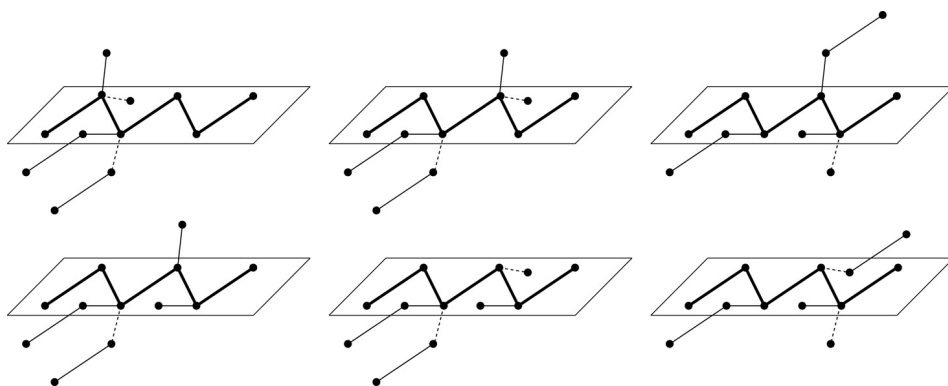| $p = 2$ $q = 2$	Automorphism groups								
n	S_1	S_2	$(S_2)^2$	S_3	$(S_2)^3$	$S_2 \times S_3$	$(S_2)^4$	$(S_3)^2$	Sum
11	0	0	0	0	0	0	1	1	2
12	0	0	6	0	5	6	0	1	18
13	0	26	36	8	5	12	2	1	90
14	16	153	77	24	13	18	0	1	302
15	140	403	155	48	13	24	2	1	786
16	456	902	232	80	21	30	0	1	1722
17	1184	1616	358	120	21	36	2	1	3338
18	2472	2727	471	168	29	42	0	1	5910
19	4668	4145	645	224	29	48	2	1	9762
20	7984	6108	794	288	37	54	0	1	15266
21	12896	8470	1016	360	37	60	2	1	22842
22	19680	11525	1201	440	45	66	0	1	32958
23	28940	15071	1471	528	45	72	2	1	46130
24	41016	19458	1692	624	53	78	0	1	62922
25	56640	24428	2010	728	53	84	2	1	83946

Fig. 9. Graph families as a function of the number of vertices.

Fig. 10. 6 stereo-isomers of $C_{12}H_{26}$, 2 methyls and 2 ethyls, automorphism group $S_2 \times S_3$.

Besides, this exhaustive approach can be adapted to specific graph subfamilies of alkane molecules which are mostly encountered in practical applications. It is worth noting that the enumeration of stereo-isomers of such subfamilies with the holistic approach of generating

Table 5

CPU time in minutes for the generation of the structural isomers as a function of n on a Pentium III, 1 GHz computer

n	11	12	13	14	15	16	17	18	19	20	21	22	23
t	0.0	0.0	0.0	0.0	0.1	0.2	0.4	1.0	2.6	7.3	23.1	99.5	511.0

Table 6

Distribution of isomers according to the automorphism group. The lateral branches are composed of 4 methyl radicals and no ethyl radical

$p = 4$ $q = 0$	Automorphism group								
n	S_1	S_2	$(S_2)^2$	S_3	$(S_2)^3$	$S_2 \times S_3$	$(S_2)^4$	$S_2 \times (S_3)^2$	Sum
8	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	2	1	1	4
10	0	0	2	1	3	4	0	1	11
11	0	2	9	3	3	6	1	1	25
12	0	11	16	6	6	8	0	1	48
13	2	25	30	10	6	10	1	1	85
14	6	53	42	15	9	12	0	1	138
15	16	89	63	21	9	14	1	1	214
16	32	146	80	28	12	16	0	1	315
17	60	214	108	36	12	18	1	1	450
18	100	310	130	45	15	20	0	1	621
19	160	420	165	55	15	22	1	1	839
20	240	565	192	66	18	24	0	1	1106
21	350	727	234	78	18	26	1	1	1435
22	490	931	266	91	21	28	0	1	1828
23	672	1155	315	105	21	30	1	1	2300
24	896	1428	352	120	24	32	0	1	2853
25	1176	1724	408	136	24	34	1	1	3504

functions is out of reach. Moreover, this new approach provides an important new feature: it becomes possible to use a combinatorial approach to characterize the three-dimensional geometry and number of the various isomers, and thus to deduce some physical or chemical properties of their mixtures. It opens new perspectives in the field of organic and statistical chemistry since it is intrinsically adequate for problems in which the spatial configurations of the various molecules have a key role in the global properties of the mixture.

Acknowledgments

References

Balaban, A.T. (Ed.), 1976. Chemical Applications of Graph Theory. Academic Press (Harcourt Brace Jovanovich Publishers), London.

Balaban, A.T., 1987. Numerical modelling of chemical structures: local graph invariants, and topological indices. In: Graph Theory and Topology in Chemistry (Athens, Ga., 1987). Elsevier, Amsterdam, pp. 159–176.

Balaban, A.T., 1993. Solved and unsolved problems in chemical graph theory. In: Quo Vadis, Graph Theory? North-Holland, Amsterdam, pp. 109–126.

Biggs, N.L., Lloyd, K.E., Wilson, R.J., 1986. Graph Theory. 1736–1936, 2nd edition. The Clarendon Press, Oxford University Press, New York.

Chaunier, C., Lygerōs, N., 1992. The number of orders with thirteen elements. Order 9 (3), 203–204.

Chaunier, C., Lygerōs, N., 1994. Le nombre de posets à isomorphie près ayant 12 éléments. Theoret. Comput. Sci. 123 (1), 89–94. (Number Theory, Combinatorics and Applications to Computer Science (Marseille, 1991)).

Harary, F., Norman, R.Z., 1960. Dissimilarity characteristic theorems for graphs. Proc. Amer. Math. Soc. 11, 332–334.

Harary, F., Palmer, E.M., 1973. Graphical Enumeration. Academic Press, New York.

Jaffe, H.H., Orchin, M., 1965. Symmetry in Chemistry. John Wiley and Sons.

Laurent, F., Massot, M., 2001. Multi-fluid modeling of laminar poly-dispersed spray flames: Origin, assumptions and comparison of sectional and sampling methods. Combust. Theory Model. 5 (4), 537–572.

Lygerōs, N., Mizony, M., 1996. Construction de posets dont le groupe d'automorphismes est isomorphe à un groupe donné. C. R. Acad. Sci. Paris Sér. I Math. 322 (3), 203–206.

Massot, M., Kumar, M., Gomez, A., Smooke, M.D., 1998. Counterflow spray diffusion flames of heptane: Computations and experiments. In: 27th Symposium (International) on Combustion. The Combustion Institute, pp. 1975–1983.

McKay, B.D., 1981. Practical graph isomorphism. In: Proceedings of the Tenth Manitoba Conference on Numerical Mathematics and Computing, Vol. I (Winnipeg, Man., 1980), vol. 30, pp. 45–87.

Pólya, G., 1937. Kombinatorische anzahlbestimmung für gruppen, graphen und chemische verbindungen. Acta. Math. 68, 145–254.

Pólya, G., Read, R.C., 1987. Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds. Springer-Verlag, New York (Pólya's contribution translated from the German by Dorothee Aeppli).

Randić, M., Guo, X., Bobst, S., 2000. Use of path matrices for a characterization of molecular structures. In: Discrete Mathematical Chemistry (New Brunswick, NJ, 1998). Amer. Math. Soc., Providence, RI, pp. 305–322.

Read, R.C., 1972. Some recent results in chemical enumeration. In: Graph Theory and Applications (Proc. Conf., Western Michigan Univ., Kalamazoo, Mich., 1972; Dedicated to the memory of J. W. T. Youngs). In: Lecture Notes in Math., vol. 303. Springer, Berlin, pp. 243–259.

Read, R.C., 1976. In: Balaban, A.T. (Ed.), The Enumeration of Acyclic Chemical Compounds. Academic Press (Harcourt Brace Jovanovich Publishers), London, pp. 25–62 (Chapter 4).