

On Morgan-trees*

István Lukovits^{a,**} and Ivan Gutman^b

^a *Chemical Research Center, H-1525 Budapest, P. O. Box 17, Hungary*

^b *Faculty of Science, University of Kragujevac, Yu-34000 Kragujevac, P. O. Box 60, Yugoslavia*

Received June 4, 2001; revised October 29, 2001; accepted October 30, 2001

Morgan-trees (MTs) are a subclass of physical trees. Because their number is much less than the number of physical trees, MTs can be used for generating acyclic isomers of alkanes and alkyl derivatives. In this paper we have summarised the properties of MTs, devised a new method to obtain the recurrence formula accounting for the number of various MTs containing N vertices, and finally derived an analytical expression for this formula. In addition the method of deleting redundant structures was discussed.

Key words: Morgan-trees, adjacency matrix, physical tree.

INTRODUCTION

The concept of a physical tree (corresponding to the structural formula of an acyclic molecule) has been first presented by Knop *et al.*¹ According to their definition a physical tree is obtained by assigning labels to the vertices consecutively and each vertex to be labeled must be adjacent to an already labeled vertex. As a result of this procedure each vertex has only a single neighbour with a lower ordinal, the only exception is vertex 1, which has no neighbor labeled with a lower ordinal. Figure 1 shows the hydrogen-suppressed graph (structural formula) of 3-ethyl-2,4-dimethylpentane and its adjacency matrix if the structure is a non-physical tree. Only the upper-right triangle of **A** has been listed, zeros, except those in the diagonal, have

* Dedicated to Professor Milan Randić on the occasion of the 70th birthday.

** Author to whom correspondence should be addressed. (E-mail: lukovits@chemres.hu)

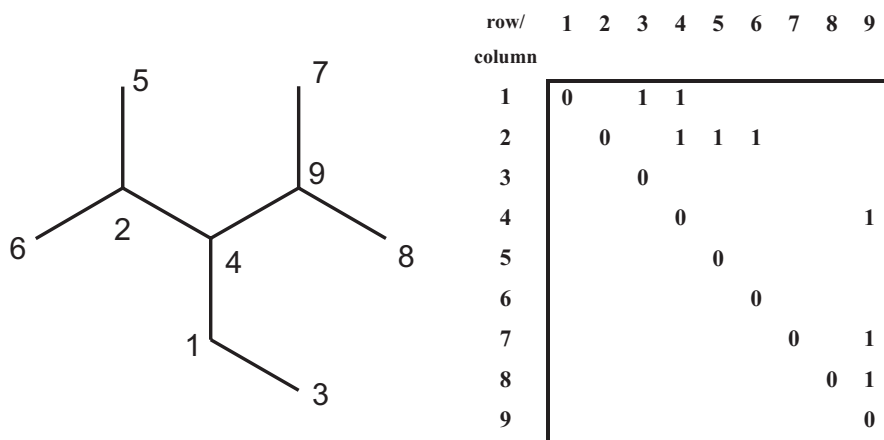


Figure 1. Hydrogen suppressed structure representing 3-ethyl-2,4-dimethylpentane and the right upper triangle of its adjacency matrix. Only zeros in the diagonal have been written out explicitly, the non-diagonal entries with zero value have been omitted for the sake of clarity.

been omitted for the sake of clarity. Figure 2 again shows 3-ethyl-2,4-dimethylpentane, the vertices of which have been numbered to obtain a physical tree. The adjacency matrix A of physical trees will contain a single non-zero entry in each column of the upper right triangular.² Observe that in case of physical trees entry $A_{1,2}$ is always equal to 1. Because of this property the generation of physical trees is easy, and it can be seen (Figure 2), that the total number of N -vertex physical trees is $(N-1)!$.

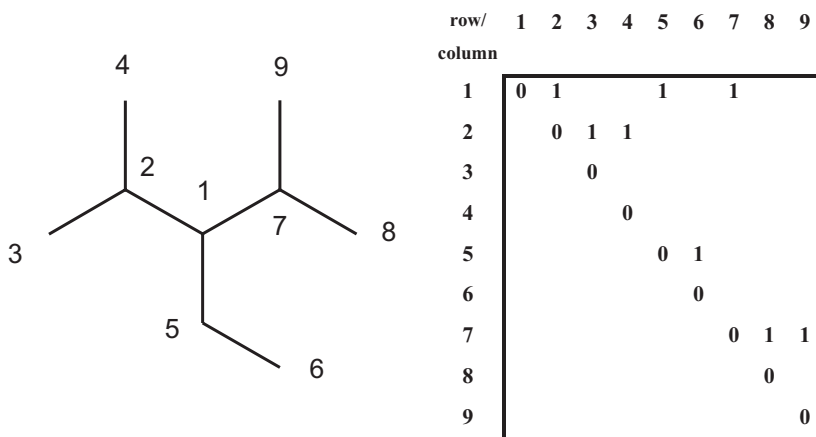


Figure 2. The structure depicted in Figure 1. The labeling corresponds to a physical tree. The compressed adjacency matrix is given in parentheses.

The number of the labeled structures can further be reduced by applying the naming algorithm of Morgan.³ The aim of this paper was to devise a method by which the number of Morgan-trees (MTs) for any N can be obtained. Morgan's algorithm has been modified in this work for reasons to be explained in the next section. The structure of the paper is the following: the concept of MT and the notion of the compressed adjacency matrix will be defined in the next section. The number of labelled structures obtained with the modified Morgan naming algorithm will be discussed in the third and fourth sections. The last section will be devoted to the selection procedure, which eliminates redundant MTs (*i.e.* identical non-labelled structures) and resulting in an exhaustive and non-redundant list of codes of structural isomers of alkanes. As an example the codes of all 802 structural isomers of alkanes containing 13 carbons have been generated.

MORGAN-TREES AND THE COMPRESSED ADJACENCY MATRIX

The concept of a Morgan-tree (MT) has been introduced by one of the present authors⁴ in 1999 and has been named in honour of H. L. Morgan, who proposed a naming algorithm,³ which is a predecessor of the variant to be examined in this paper. In Kvasnička's and Pospichal's interpretation⁵ Morgan's naming algorithm consists of the following steps:

1. Assign number 1 to a vertex of the highest degree.
2. If the actual vertex is of degree p , then assign numbers 2, 3, ..., $p + 1$ to adjacent vertices.
3. Consider next vertex possessing the lowest label and which still has non-indexed neighbors. Label its q adjacent (and non-labeled) neighbors by using the consecutive numbers $p + 2, p + 3, \dots, p + q$.
4. Repeat the same procedure for remaining vertices 3, 4, ..., $N-1$ in turn.

In this work Morgan's algorithm has been modified.⁴ The numbering starts at a vertex of the *lowest* degree (the endpoint in the case of trees) now. A MT is also a physical tree, but not every physical tree is a MT. Figure 3 displays a MT (3-ethyl-2,4-dimethylpentane). In a MT the following relations hold: if $A_{i,j} = 1$ and $A_{k,m} = 1$, and $i < k$, then $j < m$. If $j < m$, then $i \leq k$. Observe the »stair-like« structure of matrix \mathbf{A} (Figure 3).

The adjacency matrix of a physical tree (and of course, of a MT, too) can be considerably simplified by introducing the compressed adjacency matrix (CAM) denotation. CAM is a one-dimensional matrix \mathbf{C} containing $N-1$ entries and is defined in the following way:

$$\text{If } A_{i,k+1} = 1, \text{ then } C_k = i . \quad (1)$$

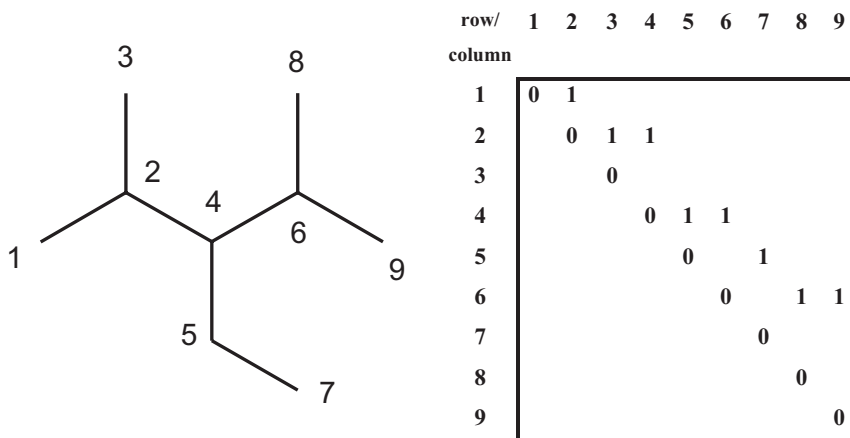


Figure 3. The structure depicted in Figure 1. The labeling corresponds to a Morgan-tree. The compressed adjacency matrix is given in parentheses.

In other words, the i -th entry of \mathbf{C} refers to the $(i + 1)$ th column of matrix \mathbf{A} , and the value of this entry is equal to the row number of the non-vanishing entry in \mathbf{A} . For example, the CAM code of the labelled structure depicted in Figure 2 is $\mathbf{C} = (1, 2, 2, 1, 5, 1, 7, 7)$. From any \mathbf{C} the corresponding adjacency matrix \mathbf{A} can be restored, and (of course) from any \mathbf{A} , the underlying graph can also be reconstructed. Each vertex i , appears in \mathbf{C} exactly $d_i - 1$ times, where d_i denotes the degree of vertex i . The only exception is again vertex 1, which appears d_1 times. Therefore endpoints (except vertex 1) do not appear in \mathbf{C} . CAMs of MTs can be obtained from the (labelled) structure directly, without resorting to the adjacency matrix.

As an example consider the structure depicted in Figure 3. The first entry of \mathbf{C} is always equal to 1. Then write down numeral 2 $d_2 - 1$ (two) times, then numeral 3 $d_3 - 1$ (zero) times, *etc.*, and finally the number $N - 1$ $d_{N-1} - 1$ times. Vertex N must be an endpoint and will not appear in the CAM. As a result we obtain for 3-ethyl-2,4-dimethylpentane (Figure 3): $\mathbf{C} = (1, 2, 2, 4, 4, 5, 6, 6)$. Any CAM denoting a MT is a *non-decreasing* sequence of integers.

Alternatively, a simple procedure can be applied to restore the MT from its underlying CAM, without resorting to the adjacency matrix. For this a single line has to be drawn bearing numbers 1 and 2 at its endpoints. Then starting from vertex 2, $d_2 - 1$ lines have to be drawn, and the endpoints are numbered from 3 through $2 + d_2 - 1 = d_2 + 1$. The same procedure has to be repeated for vertex 3, 4, *etc.* No lines emanate from endpoints.

The number of MTs is considerably less than that of physical trees and the number of physical trees in turn is just a small fraction of randomly numbered structures.

RECURRENCE RELATIONS

In order to determine the number of ways, in which vertex N can be attached to vertex k of the $N-1$ vertex MT, we introduce the (infinite) accessibility matrix S (Table I). Let entry $S_{k,N}$ denote the number of ways vertex N can be attached to vertex k , where k is a vertex of a MT consisting of $N-1$ vertices and $N-2$ edges. Observe that $S_{1,2} = 1$. There are two possibilities to attach vertex 3 to the labelled structure consisting of two vertices: it may be connected to vertex 1 or to vertex 2. Therefore $S_{1,3} = S_{2,3} = 1$. Once vertex 3 has been connected to vertex 2, vertex 4 can only be attached to vertex 2 or to vertex 3, but it can *not* be attached to vertex 1 (Figure 4). If, however, vertex 3 is adjacent to vertex 1, then vertex 4 can be connected to any of the vertices 1, 2 or 3. Observe that vertex N can only be attached to vertex 1, if vertex $N-1$ has also been attached to vertex 1. This means that all vertices $1 < k < N-1$ must also have been attached to vertex 1. Therefore for any N , $S_{1,N} = 1$ Table I).

TABLE I
Matrix of accessibilities

Row/ Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
2		0	1	2	3	4	5	6	7	8	9	10	11	12
3			0	2	5	9	14	20	27	35	44	54	65	77
4				0	5	14	28	48	75	110	154	208	273	350
5					0	14	42	90	165	275	429	637	910	1260
6						0	42	132	297	572	1001	1638	2548	3808
7							0	132	429	1001	2002	3640	6188	9996
8								0	429	1430	3432	7072	13260	23256
9									0	1430	4862	11934	25194	48450
10										0	4862	16796	41990	90440
11											0	16796	58786	149226
12												0	58786	208012
13													0	208012
14														0

There are two possibilities to attach vertex 4 to vertex 2 because there are two different MTs containing three vertices (Figure 4). Therefore $S_{1,3} + S_{2,3} = S_{2,4} = 2$. Similarly, the number of ways vertex 4 can be attached to ver-

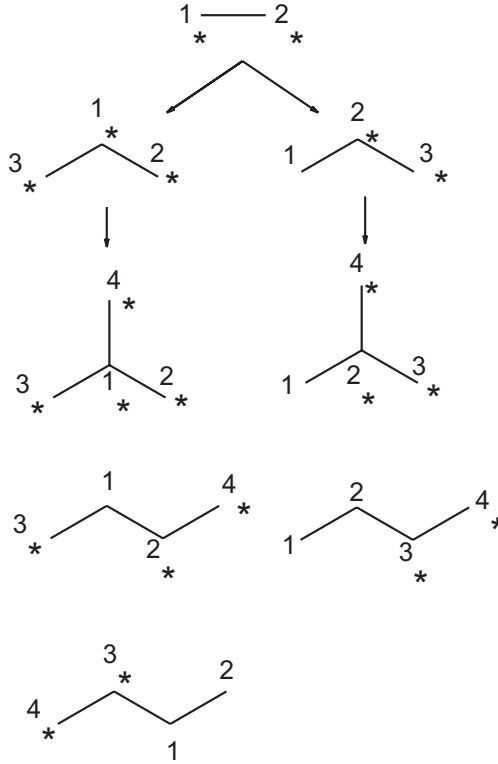


Figure 4. The evolution of a MT containing four vertices. Marked numbers indicate a vertex which the next ($N+1$ th) vertex can be attached to.

tex 3 is equal to the total number of MTs containing 3 vertices: $S_{1,3} + S_{2,3} = S_{2,4} = S_{3,4} = 2$ ($S_{2,4} = S_{3,4}$, because the number of ways vertex 4 can be attached to vertex 3 is equal to the total number of labeled 3-vertex MTs, that is to $S_{1,3} + S_{2,3}$, which in turn is equal to $S_{2,4}$. Observe also that $S_{N-2,N} = S_{N-1,N}$, for $N \geq 3$). Therefore, the total number of all four vertex MTs is $S_{1,4} + S_{2,4} + S_{3,4} = 1 + 2 + 2 = 5$, which in turn is equal to the number of ways vertex 5 can be attached to vertex 4 (Figure 4). In general $S_{i,N}$ is the number of possibilities an $N-1$ vertex MT can be obtained by attaching vertex $N-1$ to any vertex $j \leq i$. Alternatively equation $S_{2,4} = S_{2,3} + S_{1,4}$, can easily be restated in terms of a general rule. $S_{i,N-1}$ is the number of MTs obtained by attaching vertex $N-1$ to vertex i of all $N-2$ vertex MTs. $S_{i-1,N}$ is the sum of ways vertex $N-1$ can be attached to vertex 1, 2, ..., $i-1$, therefore $S_{i,N}$ is the number of ways vertex $N-1$ can be attached to vertex i , that is $S_{i,N-1} + S_{i-1,N}$.

Therefore

$$S_{i,N-1} + S_{i-1,N} = S_{i,N} . \tag{2}$$

Eq. (2) is a recursion scheme which can be used to generate all entries of matrix S . $S_{N-1,N}$ denotes the number of $N-1$ vertex MTs. The actual values of the entries of matrix S , from $N = 1$ through $N = 14$, are listed in Table I.

DERIVATION OF THE ANALYTICAL FORMULA

The elements of the first row in Table I are all equal to unity; $S_{1,j} = 1$. Evidently, the j -th element of the second row is equal to $S_{2,j} = j-2$.

Now, the difference between the j -th and the $(j-1)$ -th element of the third row of Table I is just the j -th element of the second row, *i.e.*, $j-2$. Therefore the elements of the third row of Table I must be given by a quadratic polynomial in the variable j . Direct calculation gives that this polynomial is $(j^2-3j)/2$ which we may write also in the form

$$S_{3,j} = j(j-3)/2.$$

The difference between the j -th and the $(j-1)$ -th element of the fourth row of Table I is the j -th element of the third row, *i.e.*, $j(j-3)/2$. Therefore the elements of the fourth row of Table I must satisfy a third-degree polynomial in the variable j . Direct calculation yields $(j^3-3j^2-4j)/6$ which we may write also as

$$S_{4,j} = j(j+1)(j-4)/6.$$

Analogous reasoning gives for the fifth row of Table I:

$$S_{5,j} = j(j+1)(j+2)(j-5)/24$$

whereas for the sixth row:

$$S_{6,j} = j(j+1)(j+2)(j+3)(j-6)/120$$

The general pattern of these expressions is now evident: the j -th element of the i -th row of Table I is given by

$$S_{i,j} = j(j+1)(j+2) \dots (j+i-3)(j-i)/(i-1)!$$

or, in another form:

$$S_{i,j} = \frac{(j-i)(j+i-3)!}{(j-1)!(i-1)!} \tag{3}$$

After the analytical form of the expression for $S(i,j)$ has been guessed (as outlined above), a formal proof of its general validity is easily achieved by mathematical induction.

Formula (3) holds for both $j = i$ and $j > i$, including the case $i = 1$. It also satisfies equation (2).

From Eqs. (2) and (3) we obtain the number of all MTs containing N vertices:

$$S_{N,N+1} = \frac{(2N-2)!}{(N)!(N-1)!} \quad (4)$$

GENERATION OF A NON-REDUNDANT SET OF N -TREES

Before starting this section it has to be emphasized that the vertices of any tree can be numbered in such a way that the resulting (labelled) structure is a MT. The problem of generating formulas of isomers may be significantly simplified by the application of MTs, because the number of redundant structures is much less than in the general case or in the case of physical trees.⁶

As an example let's investigate the number of labelled (acyclic) trees containing $N = 13$ vertices. Due to a graph theoretical result this is equal to $13^{11} = 1792159794037$ (This figure includes the number of all possible, labelled, connected, and disconnected trees.). The number of physical trees (see Introduction) is $12! = 479001600$ and the number of Morgan-trees (Eq. (4)) is 208012.

The generation of CAMs corresponding to MTs is equivalent to the generation of a string of numbers starting with 1 in position 1 and containing $N-1$ entries, such that $C_i \leq C_{i+1}$ (see Introduction). In case of $N = 13$ the first CAM to be generated would be $C_1 = (1,1,1,1,1,1,1,1,1,1,1)$, representing a star. The same graph, however, might be represented as $C_2 = (1,2,2,2,2,2,2,2,2,2,2)$. C_2 is greater than C_1 , because the second entry in C_2 is greater than $C_1(2)$. Out of the many possibilities to label a MT we are interested in the maximal CAM and a theorem³ (called the »lowest degrees first« – LDF – rule) ensures that by observing the LDF rule the resulting CAM will be maximal. LDF trees are obtained if in addition to the rules of the Morgan naming algorithm (applied according to our modification) an additional rule is also obeyed: the vertex with the lowest degree should be labeled first. The LDF code of 3-ethyl-2,4-dimethylpentane (Figure 5) is $C = (1,2,3,3,4,4,5,5)$, and this code is greater than the corresponding code encoding the structure depicted in Figure 3. There is no maximal CAM, C_x such that $C_1 < C_x < C_2$, because the numbering starts with an endpoint. Therefore the generation

procedure can be started with C_2 . Starting the numbering at an endpoint reduces the number of possible structures to 58786.

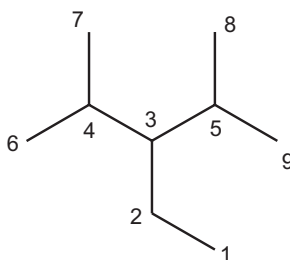


Figure 5. The numbering of 3-ethyl-2,4-dimethylpentane obtained by using the »lowest degrees first« algorithm.

Further rules⁶ ensure that $C_3 = (1,2,2,2,5,6,6,6,9,10,10,10)$ is the only maximal CAM between C_2 and $C_4 = (1,2,2,2,5,6,7,7,7,10,10,10)$. From practical point of view the generation procedure starts with the latter code (Figure 6), leaving us with 45958 labelled structures to be inspected. This number comprises only a small fraction (0.000002564%) of the original figure, 13^{11} , and comprises a mere 0.000096% of the number of physical trees. From this number we have to delete all redundant, labelled structures to retain the 802 isomers of alkanes containing 13 carbons.

Requiring that all CAMs with $C_i = C_{i+1}$ and $d_{i+1} = d_{i+2}$ be deleted, the number of remaining trees is 6132. If all CAMs corresponding to a tree, where the numbering did not start at the longest side-chain are deleted, the number of trees to be further investigated is just 1872. This set still contains all trees containing vertices of degrees greater than four. Deleting also these and applying two additional⁷ theorems, we finally obtain all 802 isomers (Figure 6). A similar list containing the graphs of all acyclic alkanes composed of 12 carbons has been published by Knop *et al.*¹

DISCUSSION

The technique based on MTs greatly reduces the number of labelled structures, which have to be checked for canonicity. Note that no comparison with other trees is required and the whole screening procedure can be performed in polynomial time.^{6,7}

Since the number of isomers increases exponentially in terms of the number of vertices, the time needed to generate and print out the formulas of isomers will also increase exponentially in terms of N , even if an (ideal-

123345578899	1233457891010	12335567991011	1233566778812	12335678991011
12334557881011	123345789101111	12335567991011	1233566788812	12335678991111
12334557881111	123355667788	123355679101010	1233566788899	123356789101010
123345578101010	1233556677888	123355679101111	123356678101011	123356789101012
123345578101111	1233556678812	1233557778910	12335667881111	123356789101111
12334558881011	12335566781111	123355779101010	123356678101010	123344455666777
12334558881111	123355777888	123355777101010	123356678101012	123344455666778
12334558891010	1233557778812	123355777101012	123356678101111	123344455666779
123345588101010	123355788812	123355777101111	12335668881011	123344455667712
123345588101111	12335578899	123355778899	1233566881011	123344455677712
123345589912	1233557881011	123355788999	123356688101010	12334445567789
12334558991111	1233557881111	123355789101010	123356688101010	12334445567789
123345589101010	12335578101010	12335577891111	123356688101012	12334445567799
123345589101111	12335578101111	123355779912	123356688101111	123344455677911
123345666777	1233558881011	1233557991011	123356689912	1233444566771011
123345667712	1233558881111	12335577991111	12335668991010	1233444566771011
123345667788	123355888101010	123355779101010	123356689101010	1233444566771010
12334567788	12335588101111	123355788101010	123356689101111	1233444567101012
12334567781111	12335589912	12335578881111	1233567778910	1233444567101111
123345678899	1233558991011	12335578891010	1233567791010	12334446666777
1233456778910	1233558991111	123355788101010	123356777101010	12334446667779
12334577791010	12335589101010	123355788101012	123356777101012	12334446667779
123345777101010	12335589101111	123355788101111	123356777101111	12334446667712
123345777101111	123356666777	123355789912	123356778899	123344466677911
123345778899	12335667779	12335578991011	123356778999	123344466677911
12334577891111	123356677712	12335578991111	1233567789912	1233444666771111
1233457799912	12335667791010	123355789101010	12335677891111	1233444677788
123345779991010	1233566771011	123355789101111	1233567799912	12334446777811
12334577991011	123356677111	1233566677899	1233567991010	12334446777811
12334577991111	12335677788	123356668899	12335677991111	123344467771111
123345779101010	1233567771011	123356668899	123356799101010	1233444677888
123345779101111	1233567771111	123356668999	123356799101012	12334446778810
12334578881011	12335677888	1233566689911	123356799101111	12334446778812
12334578881111	123356778810	1233566699912	12335678881011	12334446778910
12334578891010	12335667781111	123356669991010	123356788101010	123344467781010
123345788101010	12335667791010	12335666991111	123356789101010	123344467781111
123345788101111	123356677101010	123356669101010	123356788101010	123344467791010
1233457899912	123356677101111	123356669101012	123356788101012	1233444677101012
12334578991011	123356678899	123356669101111	123356788101111	1233444677101111
12334578991111	1233566799912	1233566677888	1233567899912	1233444667710111

ised) »zero- time« algorithm could be used to generate the structures. This argument remains valid, in case the structures are generated directly,⁸ and no redundant structures have to be deleted. The advantage of the present approach is that it might be easily extended to generate cycle-containing structures. This possibility will be investigated in a forthcoming work.

REFERENCES

1. J. V. Knop, W. R. Müller, K. Szymanski, H. W. Kroto, and N. Trinajstić, *J. Comput. Chem.* **8** (1987) 549–554.
2. I. Gutman, W. Linert, I. Lukovits, and A. A. Dobrynin, *J. Chem. Inf. Comput. Sci.* **37** (1997) 349–354.
3. H. L. Morgan, *J. Chem. Doc.* **5** (1965) 107–113.
4. I. Lukovits, *J. Chem. Inf. Comput. Sci.* **39** (1999) 563–568.
5. V. Kvasnička and J. Pospichal, *J. Chem. Inf. Comput. Sci.* **30** (1990) 99–105.
6. I. Lukovits, *J. Chem. Inf. Comput. Sci.* **40** (2000) 361–366.
7. I. Lukovits, in: D. Rouvray and R. B. King (Eds), *Chemical Graph Theory* (Proceedings of the Harry Wiener Memorial Conference), Athens, Ga, Horwood Publishing Ltd., Westergate, Chichester, 2001, pp. 327–337.
8. R. C. Read, *Annals of Discrete Math.* **2** (1978) 107–120.

SAŽETAK

O Morganovim stablima

Istvan Lukovits i Ivan Gutman

Morganova stabla podklasa su fizikalnih stabala. Budući da je njihov broj mnogo manji nego broj fizikalnih stabala, Morganova stabla mogu se rabiti za generiranje izomera alkana i alkilnih derivata. U radu su prikazana svojstva Morganovih stabala, izvedena je nova rekurzivna formula za prebrojavanje različitih Morganovih stabala koja sadrže N vrhova te je dobiven analitički izraz za tu formulu. Također se raspravlja i o metodi kojom se uklanjaju suvišne strukture.