



Home Search Collections Journals About Contact us My IOPscience

Enumerating tree-like polyphenyl isomers

This content has been downloaded from IOPscience. Please scroll down to see the full text.

J. Stat. Mech. (2012) P12018

(http://iopscience.iop.org/1742-5468/2012/12/P12018)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 59.77.43.191 This content was downloaded on 12/07/2015 at 14:18

Please note that terms and conditions apply.

ournal of Statistical Mechanics: Theory and Experiment

# Enumerating tree-like polyphenyl isomers

## Kecai Deng, Jianguo Qian<sup>1</sup> and Fuji Zhang

School of Mathematical Sciences, Xiamen University, Xiamen 361005, Fujian, People's Republic of China E-mail: kecaideng@126.com, jgqian@xmu.edu.cn and fjzhang@xmu.edu.cn

Received 3 September 2012 Accepted 26 November 2012 Published 20 December 2012

Online at stacks.iop.org/JSTAT/2012/P12018 doi:10.1088/1742-5468/2012/12/P12018

**Abstract.** Enumeration of molecules is one of the fundamental problems in bioinformatics and plays an important role in drug discovery, experimental structure elucidation (e.g., by using NMR or mass spectrometry), molecular design and virtual library construction. We consider the enumeration of tree-like polyphenyls ( $C_{6n}H_{4n+2}$ ). For this purpose, we define two generating functions T(x) and R(x) involving the numbers  $t_n$  and  $r_n$  of tree-like polyphenyls (TLpolyphenyls) and monosubstituted tree-like polyphenyls (MTL-polyphenyls), respectively. By characterizing the symmetry groups with respect to TLpolyphenyls and MTL-polyphenyls, we establish two functional equations for these two generating functions. This yields for the first time an efficient recursion formula for calculating the numbers  $t_n$  and  $r_n$ . The two functional equations are also the fundamentals for analyzing their asymptotic behaviors, from which we derive the precise asymptotic values for both  $r_n$  and  $t_n$ . The resulting asymptotic values are shown to fit well to the numerical results obtained by using our recursion formula. Finally, we give an explicit enumerating expression for TLpolyphenyls of a particular type: the linear polyphenyls.

**Keywords:** exact results, bioinformatics, computational biology

<sup>&</sup>lt;sup>1</sup> Author to whom any correspondence should be addressed.

## Contents

1.	Introduction	<b>2</b>		
2.	Results and discussion			
	2.1. Terminology and definitions	4		
	2.2. The recursion counting formula for monosubstituted tree-like polyphenyls .	4		
	2.3. The recursion counting formula for tree-like polyphenyls	6		
	2.4. The explicit enumerating expression for linear polyphenyls	9		
	2.5. Asymptotic analysis	10		
3.	Conclusions and numerical results	11		
Acknowledgments				
References				

## 1. Introduction

Enumeration of molecules is one of the fundamental problems in bioinformatics and chemoinformatics, and has attracted the interest of chemists, biologists and mathematicians for more than a century [1, 2]. It is also important from a practical viewpoint because it plays an important role in drug discovery, experimental structure elucidation (e.g., by using NMR or MS), molecular design [3], virtual library construction, hypothesis testing and experiment optimization [4].

We consider the problem of tree-like polyphenyl enumeration. Tree-like polyphenyls are tree-like (linear or branched) polymers with chemical formula  $C_{6n}H_{4n+2}$  consisting of benzene rings any two of which are connected by at most one single bond. In practice, tree-like polyphenyls are starting materials for producing all kinds of derivatives which are widely used in organic synthesis, drug synthesis, heat exchangers, etc [5]. In particular, the tree-like polyphenyl of lowest weight, i.e., biphenyl, which forms colorless crystals and occurs naturally in coal tar, crude oil and natural gas, is the parent compound of the polychlorinated biphenyls (PCBs) and, therefore, extensively applied in printing and dyeing [6, 7]. Biphenyl is also an intermediate for the production of a host of some emulsifiers, optical brighteners, crop protection products and plastics [8].

In 1960s, Kovacic *et al* synthesized a series of *p*-polyphenyls by using molybdenum pentachloride [9]. Since then, a large number of various tree-like (linear or branched) polyphenyls and their derivatives have been successively synthesized in the past fifty years, e.g., we refer the reader to [10]–[13] for details. Recently, by taking advantage of modern developments in aryl–aryl cross-coupling reactions, Mathew *et al* established a unified synthetic approach that allows the stepwise synthesis of oligomers of arbitrary length and synthesized *o*-phenylene oligomers up to the octamer [14].

In addition to the interest in the synthesis, chemists also paid attention to studying their chemical and physical properties such as the zero-field splitting parameters [15], structural features [16], electrical conductivity response [17] and thermodynamic properties [18, 19] etc. As regards mathematical aspects, the Wiener index [20, 21], the edge-Wiener index [22], the Merrifield–Simmons index and Hosoya index [23], k-matchings and k-independent sets [24] and the degree distance index [25] for their chemical graphs were studied.

Enumerating tree-like molecules or acyclic chemical compounds may date from 1857 when Cayley [1] successfully found a recursive formula for the number of trees or rooted trees. In 1931, Henze and Blair [26] improved Cayley's method and established a recursive procedure for calculating the number of quartic trees, i.e., the number of branches at each vertex is at most 4, which would be the first solution to the counting problem for alkane isomers in chemistry [27]. In 1937, Pólya [2] developed a powerful theory for treating the symmetries of certain kinds of configurations under a given permutation group, which nowadays is known as Pólya's theorem or the Redfield–Pólya theorem and represents one of the cornerstones of modern combinatorics. Using this theory, Pólya obtained the recursion formulas of Cayley, Henze and Blair, and was the first to develop the technique for analyzing the asymptotic behaviors for the numbers of trees and rooted trees. In Pólya's paper, purely combinatorial methods are employed for the developing of relations between the generating functions in which trees are treated as centered and bicentered.

The more perfectly improved technique for counting trees was introduced by Otter [27] in 1948. This technique is based on his dissimilarity characteristic theorem, by means of which the generating function for trees can be exactly expressed in terms of that for rooted trees and, therefore, the methods used in the past are simplified both theoretically and practically. On the basis of his theory, Otter also established a general method of analysis which yields asymptotic values for the coefficients involved in each of the generating functions and, moreover, the trees are not necessarily treated as centered or bicentered.

In general, it is difficult to get an explicit expression for the number of tree-like configurations. Instead, finding a recursion counting formula to deal with the problem has been shown to be an effective approach. For this reason, when a recursion formula is established, its asymptotic behavior is analyzed as well. Pólya and Otter's methodology for analyzing the asymptotic behavior was systematically summarized by Harary *et al* [28], as the so-called 'twenty-step algorithm'.

Following the enumerating theories and techniques mentioned above, lots of treelike molecules beyond alkanes have also been discussed, e.g.: tree-like polyhexes (with asymptotic analysis) [29]; phenylenes [30]; tree-like polymers [31]; radicals, monoalcohols, glycols and esters [32]; aliphatic cyclopropane derivatives [33]; etc. As regards mathematical aspects, a large number of trees or tree-like graphs with some specific requirements were also discussed; we refer the reader to, for example, [34]–[42] and the references cited therein.

We note that in the above enumeration problems, all the molecules are modeled, in terms of graph theory, as unlabeled trees. In practice, however, some enumeration problems in other areas are often modeled as those of certain kinds of labeled trees; we refer the reader to, for example, Friedberg's work [43].

To deal with the problem of tree-like polyphenyl (TL-polyphenyl) enumeration, we define two generating functions R(x) and T(x): i.e., two power series whose coefficients represent the numbers  $r_n$  and  $t_n$  of monosubstituted tree-like polyphenyls (MTL-polyphenyls) and TL-polyphenyls with n benzene rings, respectively. We determine the symmetry groups arising from the symmetry around the benzene rings and the single



**Figure 1.** (a) A tree-like polyphenyl (left) and its contracted tree (right); (b) a monosubstituted tree-like polyphenyl (left) and its contracted rooted tree (right), where the black vertex represents the root benzene ring; (c) the general structure of a monosubstituted tree-like polyphenyl in which some  $T_i$  may be empty.

bonds connecting two benzene rings. Using these specified symmetry groups and Otter's dissimilarity characteristic theorem, we establish two functional equations for these two generating functions. This yields for the first time an efficient recursion formula for calculating the numbers  $r_n$  and  $t_n$ , by which the numerical results for n up to 100 are tabulated as examples. The above two functional equations are also the fundamentals of the 'twenty-step algorithm' for analyzing the corresponding asymptotic behaviors, by means of which we derive the precise asymptotic values for both  $r_n$  and  $t_n$ . The asymptotic values are shown to fit well to the numerical results obtained by using our recursion formulas. Finally, we give an explicit enumerating expression for TL-polyphenyls of a particular type: the linear polyphenyls.

## 2. Results and discussion

## 2.1. Terminology and definitions

In the graph theoretical language, the molecular graph (i.e., the graph representing the skeleton of its carbon atoms and C–C bonds) of a TL-polyphenyl is modeled as a hexagonal system in which any two hexagons are connected by at most one edge (which represents a single bond), and its *contracted graph*, obtained from the molecular graph by contracting each hexagon to a vertex, is a tree, called the *contracted tree*; see figure 1(a). In particular, if the contracted tree is a path, then it is called a *linear polyphenyl*.

A monosubstituted tree-like polyphenyl (MTL-polyphenyl,  $C_{6n}H_{4n+1}Z$ ) is obtained from a TL-polyphenyl by replacing an H atom by a substituted ligand Z (e.g., a chlorine atom). We call this C–Z bond and the corresponding substituted benzene ring the *root* bond and the root ring, respectively. Consequently, we call the resulting contracted tree, the corresponding edge and the corresponding vertex the *rooted tree*, the *root edge* and the *root vertex*, respectively; see figure 1(b).

## 2.2. The recursion counting formula for monosubstituted tree-like polyphenyls

It would be more helpful to consider an MTL-polyphenyl with n benzene rings as one constructed by fusing a C-H bond of a 'smaller' TL-polyphenyl or nothing to each of the five numbered C-H bonds of the root ring such that the total number of benzene rings is equal to n; see figure 1(c). We call these smaller TL-polyphenyls the branches of the root ring. Since each branch has a specified C-H bond to be fused to the root ring, we may treat this branch as an MTL-polyphenyl with this C-H bond as its root bond. In

this way, a large MTL-polyphenyl can be constructed recursively, which provides us with an effective approach for determining the number of MTL-polyphenyls.

Let  $T_1, T_2, \ldots, T_5$  be five branches (some of which may be empty). We denote by  $(T_1, T_2, \ldots, T_5)$  the MTL-polyphenyl with  $T_i$   $(i \in \{1, 2, \ldots, 5\})$  as the *i*th branch of its root ring. We note that the MTL-polyphenyl obtained from  $(T_1, T_2, \ldots, T_5)$  by twisting the root bond of a branch by 180° represents the same MTL-polyphenyl. This means that, for any permutation  $\alpha$  on  $\{1, 2, \ldots, 5\}$ , two MTL-polyphenyls  $(T_1, T_2, \ldots, T_5)$  and  $(T_{\alpha(1)}, T_{\alpha(2)}, \ldots, T_{\alpha(5)})$  represent the same one if and only if  $\alpha \in G_1 =$  $\{(1)(2)(3)(4)(5), (15)(24)(3)\}.$ 

Before continuing our discussion, let us recall some elementary concepts of the classic Pólya and Burnside enumeration theories. For a permutation g of a permutation group G on an m-element set S, it is well known that g can be split into cycles in a unique way, say  $b_1$  cycles of length 1,  $b_2$  cycles of length 2,..., and  $b_m$  cycles of length m  $(m = b_1 + 2b_2 + \cdots + mb_m)$ . Then we form the product  $x_1^{b_1}x_2^{b_2}\cdots x_m^{b_m}$  and denote it by  $p_g(x_1, x_2, \ldots, x_m)$ . The cycle index of G is therefore defined by

$$P_G(x_1, x_2, \dots, x_m) = \frac{1}{|G|} \sum_{g \in G} x_1^{b_1} x_2^{b_2} \cdots x_m^{b_m} = \frac{1}{|G|} \sum_{g \in G} p_g(x_1, x_2, \dots, x_m).$$
(1)

A coloring C of S with color set C is an assignment to each element  $s \in S$  of a color  $C(s) \in C$ . Two colorings C and C' are said to be equivalent if there is a permutation  $g \in G$  such that C(g(s)) = C'(s) for any  $s \in S$ . By Burnside's enumeration theory, the number of equivalent coloring classes of S is given by

$$\frac{1}{|G|} \sum_{g \in G} \Psi(g)$$

where  $\Psi(g)$  is the number of the colorings left fixed by g [44]. In particular, if each element in S can use any color without any constraint, then  $\Psi(g) = p_g(t, t, \ldots, t)$ , where t = |C|is the number of colors.

To study the enumerating problem for MTL-polyphenyls, we need to impose some constraints on the colorings: each color c has a weight w(c) and the weight of a coloring C is defined as the sum of the weights of all the colors of  $s \in S$  assigned by C, i.e.,  $w(C) = \sum_{s \in S} w(C(s))$ . Thus, the number of equivalent coloring classes with weight k is given by

$$\frac{1}{|G|} \sum_{g \in G} \Psi_k(g),\tag{2}$$

where  $\Psi_k(g)$  is the number of the colorings with weight k and left fixed by g:

$$\Psi_k(g) = |\{\mathcal{C} : w(\mathcal{C}) = k, \mathcal{C}(g(s)) = \mathcal{C}(s) \text{ for any } s \in S\}|.$$

Denote by  $r_n$  the number of MTL-polyphenyls with n benzene rings. It is obvious that  $r_1 = 1, r_2 = 3$ . Let R(x) be the generating function of  $r_n$  in x, i.e.,

$$R(x) = \sum_{n=0}^{\infty} r_n x^n,$$

where we set  $r_0 = 1$  for the sake of recursion.

Enumerating tree-like polyphenyl isomers

## Theorem 1.

$$R(x) = 1 + xP_{G_1}(R(x), R(x^2), \dots, R(x^m)) = 1 + \frac{x}{2}[R^5(x) + R(x)R^2(x^2)].$$
(3)

**Proof.** We treat each MTL-polyphenyl  $(T_1, T_2, T_3, T_4, T_5)$  as a coloring of the root benzene ring defined as follows: the *i*th of its five C–H bonds is assigned color  $T_i$ , i = 1, 2, ..., 5. Furthermore, we define the weight of  $T_i$  as the number of benzene rings in  $T_i$ . Thus, the number of MTL-polyphenyls with n benzene rings equals the number of equivalent coloring classes with weight n - 1, i.e.,

$$r_n = \frac{1}{|G_1|} \sum_{g \in G_1} \Psi_{n-1}(g).$$
(4)

On the other hand, recalling that  $T_i$  can be regarded as an MTL-polyphenyl, the number of colors with weight k is exactly  $r_k$ , i.e., the number of MTL-polyphenyls with k benzene rings. Thus,  $\Psi_{n-1}(g)$  in (4) is exactly the coefficient of  $x^{n-1}$  in

$$R^{b_1}(x)R^{b_2}(x^2)\cdots R^{b_m}(x^m) = p_g(R(x), R(x^2), \dots, R(x^m))$$

and, therefore,  $r_n$  equals the coefficient of  $x^{n-1}$  in

$$\frac{1}{|G_1|} \sum_{g \in G_1} p_g(R(x), R(x^2), \dots, R(x^m)) = P_{G_1}(R(x), R(x^2), \dots, R(x^m)).$$

In other words,

$$R(x) = 1 + x P_{G_1}(R(x), R(x^2), \dots, R(x^m)).$$

On the other hand, note that the cycle index of  $G_1$  is

$$P_{G_1}(x_1, x_2, \dots, x_5) = \frac{1}{2}(x_1^5 + x_1 x_2^2)$$

as desired. This completes our proof.

#### 2.3. The recursion counting formula for tree-like polyphenyls

We denote by  $t_n$  the number of TL-polyphenyls (tree-like polyphenyls) with n benzene rings and by T(x) the generating function of  $t_n$  in x, i.e.,

$$T(x) = \sum_{n=1}^{\infty} t_n x^n$$

The connection between the generating functions for trees and 'rooted' trees suggests an effective way of dealing with the enumeration problem of trees, which has been extensively used for various species of trees. In this section we will establish the connection between R(x) and T(x) by applying Otter's dissimilarity characteristic theorem given below, which will also be helpful in our further discussion for analyzing the asymptotic behavior of the number  $t_n$ .

Two vertices or two edges in a tree T are called *similar* if there is an automorphism of T which takes one to the other. An edge e is called *symmetric* if its two branches (the two rooted trees of T - e with the two end vertices of e as their roots) are isomorphic; see figure 2(c). We denote by v(T) and e(T) the number of nonsimilar vertices and nonsimilar edges (except the possible symmetric edge) in T, respectively.





**Figure 2.** (a) A benzene ring and its six branches in a tree-like polyphenyl; (b) the corresponding vertex and six branches in the contracted tree; (c) an edge and its two branches.

**Theorem 2** (Theorem on page 588 of Otter [27]). In any tree T, the number of nonsimilar vertices minus the number of nonsimilar edges (except the possible symmetric edge) is equal to 1, i.e.,

$$v(T) - e(T) = 1.$$
 (5)

To apply Otter's approach, it is convenient to consider a TL-polyphenyl as its contracted tree (see section 2.2). Otter's theorem originally treats the normal trees where the branch of each vertex is 'freely interconvertible', i.e., the branch of each vertex x is interconvertible under the symmetry group  $S_d$ , where d is the degree of x. Equivalently, for the branches  $T_1, T_2, \ldots, T_d$  of a vertex x in a tree T and a permutation  $\pi \in S_d$ , the tree obtained from T by replacing the branches  $T_1, T_2, \ldots, T_d$  by  $T_{\pi(1)}, T_{\pi(2)}, \ldots, T_{\pi(d)}$ , respectively, represents T itself. However, things are different here for the contracted tree of a TL-polyphenyl, since a vertex in a contracted tree represents a benzene ring whose branches are interconvertible only under the operation of the dihedral group  $D_6$  (with the center of the benzene ring as its symmetry point); see figure 2(a). Therefore, the notion of being 'similar' for the vertices in a contracted tree is different in meaning from that for a normal tree.

For a vertex x of a contracted tree T, we denote by  $T_1(x), T_2(x), \ldots, T_6(x)$  (some of which may be empty) the branches of x in clockwise order; see figure 2(b). Under the operation of an automorphism  $\alpha$  of T,  $T_i(x)$  will be transferred to being a branch of  $\alpha(x)$ , say  $T_{i_\alpha}(\alpha(x))$  where  $i_\alpha \in \{1, 2, \ldots, 6\}$ . We denote by  $\alpha_x$  the permutation on  $\{1, 2, \ldots, 6\}$  (with respect to x) satisfying

$$\alpha_x(i) = i_\alpha, \qquad i = 1, 2, \dots, 6.$$

We call an automorphism  $\alpha$  of a tree T the  $D_6$ -restricted automorphism if  $\alpha_x \in D_6$  for each vertex x in T. Then from the above discussion, we can now give the exact meaning of the notion of being 'similar' for the vertices in a contracted tree: two vertices x and x' in a contracted tree T are called *similar* if and only if there is a  $D_6$ -restricted automorphism  $\alpha$  such that  $\alpha(x) = x'$ .

It is easy to see that the product of any two  $D_6$ -restricted automorphisms is still  $D_6$ -restricted. This means that all the  $D_6$ -restricted automorphisms of T form a subgroup of  $\operatorname{Aut}(T)$  (the automorphism group of T). We denote this subgroup by  $\operatorname{Aut}^*(T)$ . Thus, two vertices in a contracted tree are similar if they can be transferred from one to another via an automorphism in  $\operatorname{Aut}^*(T)$  and, consequently, the similarity of two edges in a

contracted tree as well as the notions of v(T) and e(T) for a contracted tree T are defined correspondingly.

Since  $\operatorname{Aut}^*(T)$  is a group, then the vertices (edges) in a contracted tree T can be partitioned into equivalent classes under the operation of  $\operatorname{Aut}^*(T)$ , and vertices (edges) in different classes are nonsimilar. Therefore, theorem 2 is still appropriate for a contracted tree with the similarity of vertices and edges defined as above, the proof of which follows word by word from the original one (replacing 'homeomorphism' in the original proof by ' $D_6$ -restricted automorphism') and is omitted here.

Now let  $\mathcal{T}_n$  be the set of all the different contracted trees (with respect to Aut<sup>\*</sup>(T)) of order n. We denote by  $v_n$   $(n \ge 1)$  and  $e_n$   $(n \ge 1)$  the sum of the numbers of the nonsimilar vertices and nonsimilar edges (except the possible symmetric edges) over all the different contracted trees in  $\mathcal{T}_n$ , respectively:

$$v_n = \sum_{T \in \mathcal{T}_n} v(T)$$
 and  $e_n = \sum_{T \in \mathcal{T}_n} e(T).$  (6)

Combining (6) with (5), we get that

$$v_n - e_n = |\mathcal{T}_n| = t_n,\tag{7}$$

where we recall that  $t_n$  is the number of TL-polyphenyls with n rings, i.e., the number of different contracted trees. Let V(x) and E(x) be the generating functions of  $v_n$  and  $e_n$ , respectively, i.e.,

$$V(x) = \sum_{n=1}^{\infty} v_n x^n$$
 and  $E(x) = \sum_{n=1}^{\infty} e_n x^n$ .

Then we may rewrite (7) as the form

$$T(x) = V(x) - E(x).$$
 (8)

We now try to deduce the expressions of V(x) and E(x) in terms of  $R(x), R(x^2), R(x^3), \ldots$  For a contracted tree T and a vertex  $x \in T$ , we denote by  $T_x$  the rooted tree of T with root x. Note that a root here has six branches (some of which may be empty), which is different from the case for an MTL-polyphenyl. An edge rooted tree  $T_e$  with root edge e is defined analogously. We note that each of the two end branches of a root edge contains at least one benzene ring.

For a contracted tree T with k nonsimilar vertices  $x_1, x_2, \ldots, x_k$ , one can see that  $T_{x_1}, T_{x_2}, \ldots, T_{x_k}$  are pairwise different (with respect to  $\operatorname{Aut}^*(T)$ ). Furthermore, any two rooted trees of two different contracted trees are different. This implies that the number of different rooted trees of order n is exactly the number  $v_n$ . Note that two rooted trees are identical if there is a permutation  $\alpha \in D_6$  on the six branches of the root which takes one to the other. On the other hand, the cycle index of  $D_6$  is

$$P_{D_6}(x_1, x_2, \dots, x_6) = \frac{1}{12}(x_1^6 + 3x_1^2x_2^2 + 4x_2^3 + 2x_3^2 + 2x_6).$$

Like in the discussion in the proof of theorem 1, we have the following relation:

$$V(x) = xP_{D_6}(R(x), R(x^2), \dots, R(x^6))$$
  
=  $\frac{x}{12}[R^6(x) + 3R^2(x)R^2(x^2) + 4R^3(x^2) + 2R^2(x^3) + 2R(x^6)].$  (9)

For the number  $e_n$ , recall that e(T) is the number of nonsimilar edges of T excepting the possible symmetric edge. This implies that  $e_n + s_n$  is exactly the number of edge rooted trees of order n, where  $s_n$  is the number of the edge rooted trees of order n whose root edges are symmetric. On the other hand, notice that each branch of an root edge has at least one vertex. Moreover, they are freely interconvertible from each other, i.e., they are interconvertible under the operation of the symmetry group  $S_2$  (with the center of the edge as its symmetry point). Note that the cycle index of  $S_2$  is  $P_{S_2}(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2)$ . Thus, we have the following relation:

$$E(x) + (R(x^{2}) - 1) = \frac{1}{2}[(R(x) - 1)^{2} + (R(x^{2}) - 1)],$$
(10)

where we notice that  $(R(x^2) - 1)$  is regarded as the generating function of  $s_n$ .

By (8)-(10), we have the following result immediately.

## Theorem 3.

$$T(x) = \frac{x}{12} [R^{6}(x) + 3R^{2}(x)R^{2}(x^{2}) + 4R^{3}(x^{2}) + 2R^{2}(x^{3}) + 2R(x^{6})] - \frac{1}{2}R^{2}(x) + \frac{1}{2}R(x^{2}) + R(x) - 1.$$
(11)

## 2.4. The explicit enumerating expression for linear polyphenyls

In this section we will give an explicit enumerating expression for TL-polyphenyls of a particular type: the linear polyphenyls. A linear polyphenyl is called odd (resp., even) if it contains an odd (resp., even) number of benzene rings, and called monosubstituted if one of the C–H bonds in its end benzene rings is substituted by a new ligand. More specifically, we call this substituted bond and the corresponding benzene ring the root bond and the root ring, respectively.

Let  $m_n$  be the number of monosubstituted linear polyphenyls and let  $a_n$  and  $b_n$  be the numbers of odd and even linear polyphenyls with n benzene rings, respectively. It is easy to see that  $m_n = 3^{n-1}$ , since when we start from the root ring and trace along the linear polyphenyl, the (i - 1)th ring has three inequivalent candidate positions for being fused by the *i*th ring, i = 2, 3, ..., n. Furthermore, the two branches of the central ring in an odd linear polyphenyl have three unequal types of fusing, in each of which the two branches are freely interconvertible, i.e., interconvertible through the symmetry group  $S_2$ (with the center of the central ring as its symmetry point); see figure 3. Recall that the cycle index of  $S_2$  is  $P_{S_2}(x_1, x_2) = \frac{1}{2}(x_1^2 + x_2)$ . So we have

$$a_{2k+1} = 3 \times \left(\frac{1}{2}(m_k^2 + m_k)\right) = \frac{1}{2}(3^{2k-1} + 3^k).$$

Similarly, the two branches of the central bond in an even linear polyphenyl are interconvertible through the symmetry group  $S_2$ , which gives that

$$b_{2k} = \frac{1}{2}(m_k^2 + m_k) = \frac{1}{2}(3^{2(k-1)} + 3^{k-1}).$$

Thus, the number of linear polyphenyls with n benzene rings is given by

$$l_n = \begin{cases} \frac{1}{2}(3^{n-2} + 3^{(n-1)/2}), & \text{if } n \text{ is odd,} \\ \frac{1}{2}(3^{n-2} + 3^{n/2-1}), & \text{if } n \text{ is even.} \end{cases}$$



**Figure 3.** Three unequal types of fusing to form an odd linear polyphenyl, where x is the central benzene ring.

## 2.5. Asymptotic analysis

In this section, we give an asymptotic analysis for the numbers of MTL-polyphenyls and TL-polyphenyls by applying the 'twenty-step algorithm' summarized by Harary *et al* [28]. The relations (3) and (11) obtained in the previous sections are the bases for implementing the twenty steps. The key points are as follows:

- (1) Show that  $0 < \sigma < 1$ , where  $\sigma$  is the radius of convergence of R(x).
- (2) From (3) we define a function  $F(x, y) = xy^5 + [xR^2(x^2) 2]y + 2$  and prove that F(x, y) is analytic for all y and x with  $|x| < \sigma^{1/2}$ . Moreover, F(x, R(x)) = 0 for all x with  $|x| \le \sigma$ .
- (3) Show that  $F_y(\sigma, R(\sigma)) = 0$  and, consequently,  $2\sigma R^5(\sigma) = 1$ , from which we can compute  $\sigma$  recursively to any desired degree of accuracy:

 $\sigma = 0.119\,318\,643\,775\,455\,832\,594\,539\,050\,847\,225\,042\,1\cdots$ 

(4) Show that  $\sigma$  is the unique singularity of R(x) on the circle of convergence [45] and a branch point of order 2 for R(x), from which we can rewrite R(x) as the form

$$R(x) = R(\sigma) - b_1(\sigma - x)^{1/2} + b_2(\sigma - x) + b_3(\sigma - x)^{3/2} + \dots;$$
(12)

(5) From (12), the coefficient  $b_1$  can be expressed in terms of  $R(\sigma)$ ,  $R(\sigma^2)$  and  $R'(\sigma^2)$  and therefore could be estimated by replacing  $R'(\sigma^2)$  by  $R'_{100}(\sigma^2)$ , which gives that

 $b_1 \approx 1.491\,311\,088\,719\,675\,463\,3\cdots$ 

where, for a function f(x), f'(x) is the derivative of f(x).

(6) Applying a result of Pólya (lemma on page 84 of [2]), we finally get the asymptotic value of  $r_n$ :

$$r_n \sim \frac{b_1}{2} \left(\frac{\sigma}{\pi}\right)^{1/2} n^{-3/2} \sigma^{-n}, \quad \text{as } n \to \infty.$$

The discussion for  $t_n$  is analogous, and gives that

$$t_n \sim \frac{3a_3}{4} \left(\frac{\sigma^3}{\pi}\right)^{1/2} n^{-5/2} \sigma^{-n}, \qquad \text{as } n \to \infty,$$

where  $\frac{3}{4}a_3 \approx 2.3372764105425946 \cdots$ .

The detailed process of the above arguments is included in the supplementary data (available at stacks.iop.org/JSTAT/2012/P12018/mmedia).

Enumerating tree	-like poly	yphenyl i	somers
------------------	------------	-----------	--------

n	$r_n$	$ ilde{r}_n$	$t_n$	$\tilde{t}_n$
1	1	1	1	0
2	3	4	1	1
3	15	16	3	2
4	81	90	9	8
5	492	537	39	40
6	3145	3426	187	214
7	21012	22788	1046	1218
8	144541	156320	6141	7308
9	1017871	1097941	38220	45626
10	7298810	7856595	245535	293842
11	53116790	57073834	1622148	1940548
12	391288147	419803350	10943458	13084128
13	2912132714	3120286214	75155413	89769997
14	21863332200	23399635848	523839906	625116836
15	165383748136	176830003465	3698202776	4409048023
16	1259274408565	1345254835192	26399302377	31445949441
17	9643959089075	10294442809810	190291425727	226482190433
18	74235684124345	79187961145484	1383511856495	1645382178342
19	574054976392095	611968698708013	10136274995498	12046357318285
20	4457335413282002	4749045262233035	74776100157898	88808890809005
21	34738042885894512	36992531460316219	555065449754974	658832310757199
22	271640637393431750	289134998555360951	4143538912470558	4915391524724526
23	2130661325140862984	2266911800430671924	31090452713611304	36862681168161090

**Table 1.** The numerical results for the numbers  $r_n, t_n$  and their asymptotic values  $\tilde{r}_n, \tilde{t}_n$  for number *n* of benzene rings from 1 to 23.

## 3. Conclusions and numerical results

Theorem 1 gives an efficient counting formula for MTL-polyphenyls:  $r_n$  is recursively given by the coefficient of  $x^n$  in

$$1 + \frac{x}{2} [R_{n-1}^5(x) + R_{n-1}(x)R_{n-1}^2(x^2)],$$
(13)

where  $R_k(x) = \sum_{i=0}^k r_i x^i$ , k = 1, 2, ... By (13) together with theorem 3, we can also get the counting formula for the number of TL-polyphenyls in terms of that for MTL-polyphenyls.

In table 1 we list the numerical results for the numbers of MTL-polyphenyls and TL-polyphenyls, and their asymptotic values with n benzene rings, for n from 1 up to 23. To give a better understanding of their asymptotic behaviors, we also illustrate the numerical results and the asymptotic values for n up to 52 in figure 4, from which we can see that the logarithms of the asymptotic values fit well to those of our numerical results.

The numbers of MTL-polyphenyls and TL-polyphenyls with n benzene rings for n from 24 up to 100, and all the MTL-polyphenyls with  $n \leq 4$  and the TL-polyphenyls with  $n \leq 5$  are tabulated in the supplementary data (available at stacks.iop.org/JSTAT/2012/P12018/mmedia).





**Figure 4.** The horizontal axis represents the number of benzene rings and the vertical axis represents the logarithms of the numbers of TL-polyphenyls and MTL-polyphenyls.

## Acknowledgments

This research was supported by NSFC (No. 10831001). The authors would like to thank the anonymous reviewers and editors for their valuable comments and suggestions.

## References

- [1] Cayley A, On the theory of the analytical forms called trees, 1857 Phil. Mag. 13 172
- [2] Pólya G and Read R C, 1987 Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds (Berlin: Springer)
- [3] Imada T, Ota S, Nagamochi H and Akutsu T, Efficient enumeration of stereoisomers of outerplanar chemical graphs using dynamic programming, 2011 J. Chem. Inf. Modelling 51 2788
- [4] Faulon J-L, Visco D P and Roe D, Enumerating molecules, 2005 Reviews in Computational Chemistry ed K Lipkowitz, R Larter and T R Cundari (Hoboken, NJ: Wiley–VCH) p 209
- [5] Weaver W C, Simmons P B and Thompson Q E, 1979 Diphenyl and Terphenyls (Kirk-Othmer Encyclopedia of Chemical Technology) ed M Grayson (New York: Wiley) p 782
- [6] Luthe G, Jacobus J A and Robertson L W, Receptor interactions by polybrominated diphenyl ethers versus polychlorinated biphenyls: a theoretical structure-activity assessment, 2008 Environ. Toxicol. Pharmacol. 25 202
- Bures M, Pekárek V and Ocelka T, Thermochemical properties and relative stability of polychlorinated biphenyls, 2008 Environ. Toxicol. Pharmacol. 25 148
- [8] Rehmann L and Daugulis A J, Biphenyl degradation kinetics by Burkholderia xenovorans, LB400 in two-phase partitioning bioreactors, 2006 Chemosphere 63 972
- [9] Kovacic P and Lange R M, Polymerization of benzene to p-polyphenyl by molybdenum pentachloride, 1963 J. Org. Chem. 28 968
- [10] Kovacic P and Kyriakis A, Polymerization of benzene top-polyphenyl by aluminum chloride-cupric chloride, 1963 J. Am. Chem. Soc. 85 454
- [11] Ibuki E, Ozasa S and Murai K, Studies of polyphenyls and polyphenylenes. I. The syntheses and infrared and electronic spectra of several sexiphenyls, 1975 Bull. Chem. Soc. Japan. 48 1868
- [12] Ibuki E, Ozasa S, Fujioka Y, Okada M and Yanagihara Y, Polyphenyl synthesis by means of the Kharash-type Grignard cross-coupling reaction, 1982 Chem. Pharm. Bull. 30 2369
- [13] Li Q R, Yang Q, Yin H and Yang S, Analysis of by-products from improved Ullmann reaction using TOF-MS and GC-TOF-MS, 2004 J. Univ. Sci. Technol. China 34 335
- [14] Mathew S M and Hartley C S, Parent o-phenylene oligomers: synthesis, conformational behavior, and characterization, 2011 Macromolecules 44 8425

- [15] Higuchi J, Hayashi K, Yagi M and Kondo H, Conformations and zero-field splittings in the lowest excited triplet states of meta- and para-polyphenyl molecules, 2002 J. Phys. Chem A 106 8609
- [16] Gagnon E, Maris T, Arseneault P M, Maly K E and Wuest J D, Structural features in crystals of derivatives of benzene with multiple contiguous phenyl substituents, 2010 Cryst. Growth Des. 10 648
- [17] Phumman P and Sirivat A, Electrical conductivity response of poly(p-phenylene)/ZSM-5 composite, 2007 Proc. 7th IEEE Int. Conf. on Nanotechnology (Hong Kong, Aug.) p 1085
- [18] Lima C F, Rocha M A, Melo A, Gomes L R, Low J N and Santos L M, Structural and thermodynamic characterization of polyphenylbenzenes, 2011 Phys. Chem. A 115 11876
- [19] Ribeiro da Silva M A V, Santos L M and Lima L M S S, Thermodynamic study of 1,2,3-triphenylbenzene and 1,3,5-triphenylbenzene, 2010 J. Chem. Thermodyn. 42 134
- [20] Bian H and Zhang F J, Tree-like polyphenyl systems with extremal Wiener indices, 2009 Match: Commun. Math. Comput. Chem. 61 631
- [21] Deng H, Wiener indices of spiro and polyphenyl hexagonal chains, 2012 Math. Comput. Model 55 634
- [22] Dou Y, Bian H and Gao H, The polyphenyl chains with extremal edge-Wiener indices, 2010 Match: Commun. Math. Comput. Chem. 64 757
- [23] Bai Y, Zhao B and Zhao P, Extremal Merrifield-Simmons index and Hosoya index of polyphenyl chains, 2009 Match: Commun. Math. Comput. Chem. 62 649
- [24] Li S, Bian H and Zhang F J, Extremal polyphenyl chains concerning k-matchings and k-independent sets, 2010 Ars Combin. 96 97
- [25] Hao Y and Wang M, Tree-like polyphenyl chains with extremal degree distance, 2011 Dig. J. Nanomater. Biostruct. 6 739
- [26] Henze H R and Blair C M, The number of structurally isomeric alcohols of the methanol series, 1931
   J. Am. Chem. Soc. 53 3042
- [27] Otter R, The number of trees, 1948 Ann. Math. 49 583
- [28] Harary F, Robinson R W and Schwenk A J, Twenty-step algorithm for determining the asymptotic number of trees of various species, 1975 J. Aust. Math. Soc. 20 483
- [29] Harary F and Read R C, The enumeration of tree-like polyhexes, 1970 Proc. Edinburgh Math. Soc. 17 1
- [30] Gutman I, Cyvin S J and Brunvoil J, Enumeration of the isomers of phenylenes, 1994 Monatsh. Chem. 125 887
- [31] Costa M R P F N and Dias R C S, Prediction of mean square radius of gyration of tree-like polymers by a general kinetic approach, 2007 Polymer 48 1785
- [32] Read R C, 1976 The Enumeration of Acyclic Chemical Compounds (Chemical Applications of Graph Theory) ed A Balaban (New York: Academic) p 24
- [33] Yeh C, Isomer enumeration of alkenes and aliphatic cyclopropane derivatives, 1996 J. Chem. Inf. Comput. Sci. 36 854
- [34] Harary F, Prins G and Tutte W T, The number of plane trees, 1964 Indag. Math. 26 319
- [35] Harary F and Palmer E M, 1973 Graphical Enumeration (New York: Academic)
- [36] Bytautas L and Klein D J, Chemical combinatorics for alkane-isomer enumeration and more, 1998
   J. Chem. Inf. Comput. Sci. 38 1063
- [37] Brunvoll J, Cyvin S J and Cyvin B N, Enumeration of tree-like octagonal systems, 1997 J. Math. Chem. 21 193
- [38] Bergeron F, Labelle G and Leroux P, 1998 Combinatorial Species and Tree-Like Structures (Cambridge: Cambridge University Press)
- [39] Balaban A T, Kennedy J W and Quintas L, The number of alkanes having n carbons and a longest chain of length d: an application of a theorem of Pólya, 1988 J. Chem. Educ. 65 304
- [40] Gordon M and Kennedy J W, The counting and coding of trees of fixed diameter, 1975 SIAM J. Appl. Math. 28 376
- [41] Knop J V, Müller W R, Szymanski K and Trinajstić N, 1985 Computer Generation of Certain Classes of Molecules (Association of Chemists and Technologists of Croatia) Zagreb
- [42] Ishida Y, Kato Y, Zhao L, Nagamochi H and Akutsu T, Branch-and-bound algorithms for enumerating treelike chemical graphs with given path frequency using detachment-cut, 2010 J. Chem. Inf. Modelling 50 934
- [43] Friedberg R, Dual trees and resummation theorems, 1975 J. Math. Phys. 16 20
- [44] Beckenbach E F, 1964 Applied Combinatorial Mathematics (New York: Wiley)
- [45] Meir A and Moon J W, On an asymptotic method in enumeration, 1989 J. Combin. Theory Ser. A 51 77