Enumerative Combinatorics and Applications
ecajournal.haifa.ac.il

ECA **6:2** (2026) Article #S2R11
https://doi.org/10.54550/ECA2026V6S2R11

Horton-Strahler Numbers for Binary Butterfly Trees: Exact Analysis

Helmut Prodinger

Department of Mathematics, University of Stellenbosch 7602, Stellenbosch, South Africa and NITheCS (National Institute for Theoretical and Computational Sciences), South Africa
Email: hproding@sun.ac.za, warrenham33@gmail.com

Received: December 13, 2025, **Accepted**: February 3, 2026, **Published**: February 6, 2026

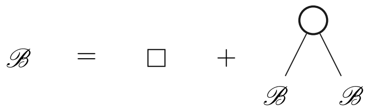
ABSTRACT: Peca suggested in a recent paper on the arxiv to consider binary butterfly trees and their Horton-Strahler numbers. The trees are obtained by glueing two binary trees together in a special way; the results are again binary trees, but with a different probability distribution. A thorough combinatorial analysis is provided and leads asymptotically to the same results as for classical binary trees.

Keywords: Binary trees; Generating functions; Horton-Strahler numbers; Mellin transform; Register function
2020 Mathematics Subject Classification: 05A15

1. Binary trees and Horton-Strahler numbers

We start with some classical observations about binary trees and Horton-Strahler numbers (also called register function in Computer science). The pioneering papers are by Flajolet, Raoult, and Vuillemin [2] and Kemp [3]. The author has collected some material in [5, 7–9].

Binary trees may be expressed by the following symbolic equation, which says that they include the empty tree and trees are recursively built from a root followed by two subtrees (left and right), which are binary trees:

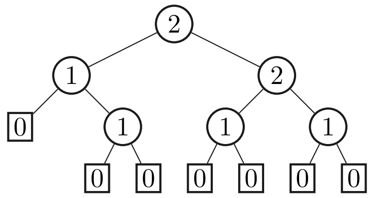


Binary trees are counted by Catalan numbers and the parameter reg, (register function, Horton-Strahler numbers) is recursively defined by attaching the number 0 to the leaves, and then working our way up: if both subtrees are labelled with the same number, the root will get 1 + this number, otherwise the larger value of the subtrees bubbles up. The value at the root is then the parameter of interest.

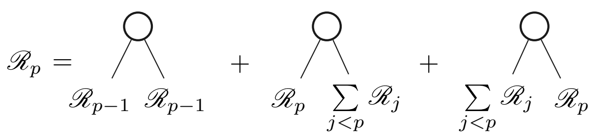
There is a recursive description of this function: $\mathsf{reg}(\square) = 0$, and if tree t has subtrees t_1 and t_2, then

$$\mathsf{reg}(t) = \begin{cases} \max\{\mathsf{reg}(t_1), \mathsf{reg}(t_2)\} & \text{if } \mathsf{reg}(t_1) \neq \mathsf{reg}(t_2), \\ 1 + \mathsf{reg}(t_1) & \text{otherwise.} \end{cases}$$

Here is an example:



Let $\mathscr{R}_p$ denote the family of binary trees with Horton-Strahler number equal to p, then one gets immediately from the recursive definition:



In terms of generating functions, these equations are translated into

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z);$$

the variable z is used to mark the size (i. e., the number of internal nodes) of the binary tree. See [7]; compare also [9].

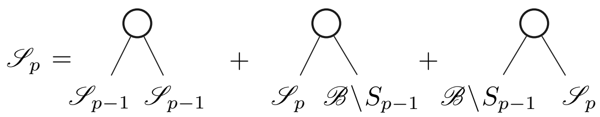
Flajolet et al. resp. Kemp were able to solve this explicitly! Nowadays, several strategies to do this are known; we only report the results as they will be used later in this paper. The substitution

$$z = \frac{u}{(1+u)^2}$$

that de Bruijn, Knuth, and Rice [1] also used, produces the nice expression

$$R_p(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}.$$

The generating function $S_p(z) = R_p + R_{p+1} + R_{p+2} + \cdots$ of binary trees with register function $\geq p$ is equally important. One can check directly that



Hence

$$S_p(z) = zS_{p-1}^2 + zS_p(B - S_{p-1}) + z(B - S_{p-1})S_p, \quad p \geq 1, \quad S_0 = B,$$

and

$$S_p(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^p}}.$$

Peca [4] introduced a "butterfly" tree by glueing two binary trees together. The motivation came from binary *search* trees, which are binary trees but with a different statistics underlying; binary search trees are equivalent to permutations and thus enumerated by $n!$, not by Catalan numbers. The interest was, however, in the Horton-Strahler numbers of the "butterfly" trees. The operation $\oplus$ is first defined on permutations: The operation $\pi_1 \oplus \pi_2$ is defined like this: Both permutations are written in one-line notation in terms of $1, 2, \ldots$; then, the second permutation is lifted up by m, if m is the size of the permutation π_1; with the resulting string of numbers, the final binary search tree is formed. A thorough analysis of the Horton-Strahler numbers in these butterfly trees was requested in [4]. This is the purpose of this paper.

The resulting tree can be described without permutations by finding the rightmost leaf of t_1, and then replacing it with the tree t_2. Peca also considered a version of $\ominus$, but that is basically a symmetric version of $\oplus$, and we will not consider it further. The resulting object is again a binary tree, but one cannot reconstruct the two binary trees t_1 and t_2 from it. In order to do so, we consider the path from the root to the rightmost leaf, and consider exactly one edge as being distinguished (marked). From this, the reconstruction is possible. We avoid considering the empty tree for t_1 since the path of interest has *no* edges.

The generating function of these marked binary trees is $A := (B-1)B = u(1+u)$. The enumeration is as follows, since B is known:

$$A = \frac{1-3z}{2z^2} - \frac{(1-z)\sqrt{1-4z}}{2z^2}, \qquad [z^n]A(z) = \frac{3(2n)!}{(n-1)!(n+2)!},\ n \geq 1.$$

When we consider the elements of $\mathscr{A}$ and ignore the marking of an edge on the rightmost path, we have a binary tree and thus can introduce the Horton-Strahler numbers on the elements of $\mathscr{A}$. No object has Horton-Strahler number equal to 0, and we introduce $T_p(z)$, the generating function of the elements of $\mathscr{A}$ with Horton-Strahler number $\geq p$. We will find a recursion for T_p that is akin to the recursion for S_p and use the fact that S_p is explicitly known. This is similar to [6] where the author solved a problem left open by Yekutieli and Mandelbrot [10], although the present situation is more involved. The further considerations deserve a section on its own.

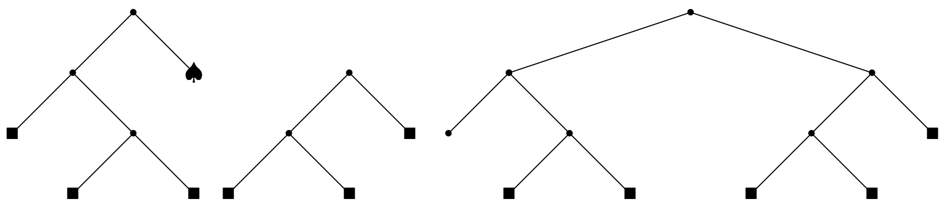


Figure 1: Two binary trees t_1, t_2, the rightmost leaf of t_1 indicated, and then t_2 glued there, with resulting binary tree $t_1 \oplus t_2$

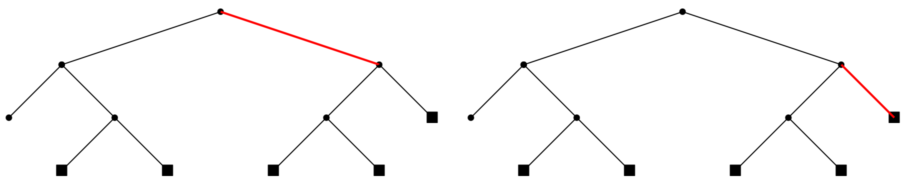


Figure 2: The resulting tree $t_1 \oplus t_2$ can originate in two possible ways from glueing two trees together.

2. Horton-Strahler numbers resulting from glueing together two binary trees

Recall that

$$S_p(z) = zS_{p-1}^2 + zS_p(B - S_{p-1}) + z(B - S_{p-1})S_p, \quad p \ge 1.$$

Using a similar reasoning, distinguishing whether the first right edge is marked (first line) or is not marked (second line), we find

$$T_p(z) = zS_{p-1}^2 + zS_p(B - S_{p-1}) + z(B - S_{p-1})S_p + zS_{p-1}T_{p-1} + zS_p(A - T_{p-1}) + z(B - S_{p-1})T_p,$$

or

$$T_p(z) = S_p + zS_{p-1}T_{p-1} + zS_p(A - T_{p-1}) + z(B - S_{p-1})T_p, \quad p \ge 2, \text{ and } T_1 = A.$$

Further

$$T_p(z)\big(1 - z(B - S_{p-1})\big) = S_p(1 + zA) + zT_{p-1}(S_{p-1} - S_p).$$

A direct computation yields

$$1 - z(B - S_{p-1}) = \frac{1 - u^{2^{p-1}+1}}{(1+u)(1 - u^{2^{p-1}})}, \quad 1 + zA = \frac{1 + u + u^2}{1 + u}.$$

Further

$$S_{p-1} - S_p = R_{p-1} = \frac{1 - u^2}{u} \frac{u^{2^{p-1}}}{1 - u^{2^p}}.$$

Some rearrangements lead to

$$T_p(z)(1 - u^{2^{p-1}+1})(1 + u^{2^{p-1}}) = \frac{1 - u^2}{u} u^{2^p}(1 + u + u^2) + T_{p-1}(1 - u)u^{2^{p-1}};$$

this first order recursion will be solved by introduction of summation factors and eventually by summation. This will be done in a few steps;

$$\frac{T_p(z)}{u^{2^p}}(1 - u^{2^{p-1}+1})(1 + u^{2^{p-1}}) = \frac{1 - u^2}{u}(1 + u + u^2) + \frac{T_{p-1}(z)}{u^{2^{p-1}}}(1 - u);$$

$$\frac{T_p(z)}{u^{2^p}}(1 - u^{2^{p-1}+1}) \prod_{j=0}^{p-1}(1 + u^{2^j}) = \frac{1 - u^2}{u}(1 + u + u^2) \prod_{j=0}^{p-2}(1 + u^{2^j}) + \frac{T_{p-1}(z)}{u^{2^{p-1}}}(1 - u) \prod_{j=0}^{p-2}(1 + u^{2^j}).$$

Evaluating the products yields

$$\frac{T_p(z)(1-u^{2^{p-1}+1})}{u^{2^p}}(1-u^{2^p}) = \frac{1-u^2}{u}(1+u+u^2)(1-u^{2^{p-1}}) + \frac{T_{p-1}(z)(1-u^{2^{p-1}})}{u^{2^{p-1}}}(1-u)$$

and

$$\frac{T_p(z)(1-u^{2^p})}{u^{2^p}(1-u)^p}(1-u^{2^{p-1}+1}) = \frac{1-u^2}{u(1-u)^p}(1+u+u^2)(1-u^{2^{p-1}}) + \frac{T_{p-1}(z)(1-u^{2^{p-1}})}{u^{2^{p-1}}(1-u)^{p-1}}$$

Introducing a further product,

$$\begin{aligned}&\frac{T_p(z)(1-u^{2^p})}{u^{2^p}(1-u)^p}\prod_{j=0}^{p-1}(1-u^{2^j+1})\\ &= \frac{1+u}{u(1-u)^{p-1}}(1+u+u^2)(1-u^{2^{p-1}})\prod_{j=0}^{p-2}(1-u^{2^j+1}) + \frac{T_{p-1}(z)(1-u^{2^{p-1}})}{u^{2^{p-1}}(1-u)^{p-1}}\prod_{j=0}^{p-2}(1-u^{2^j+1}).\end{aligned}$$

An abbreviation is useful:

$$\omega_p := \frac{T_p(z)(1-u^{2^p})}{u^{2^p}(1-u)^p}\prod_{j=0}^{p-1}(1-u^{2^j+1}),$$

then we get a form that can be summed:

$$\omega_p = \omega_{p-1} + \frac{1+u}{u(1-u)^{p-1}}(1+u+u^2)(1-u^{2^{p-1}})\prod_{j=0}^{p-2}(1-u^{2^j+1})$$

and so

$$\omega_p = \omega_1 + \sum_{h=1}^{p-1}\frac{1+u}{u(1-u)^h}(1+u+u^2)(1-u^{2^h})\prod_{j=0}^{h-1}(1-u^{2^j+1}).$$

Coming back to the original quantities $T_p(z)$,

$$\begin{aligned}T_p(z) &= \omega_p\frac{u^{2^p}(1-u)^p}{(1-u^{2^p})}\Big/\prod_{j=0}^{p-1}(1-u^{2^j+1})\\ &= \Bigg[(1+u)^2 + (1+u+u^2)\sum_{h=1}^{p-1}\frac{1-u^{2^h}}{1-u}\prod_{j=0}^{h-1}\frac{1-u^{2^j+1}}{1-u}\Bigg]\times\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^p}}\prod_{j=0}^{p-1}\frac{1-u}{1-u^{2^j+1}}.\end{aligned}$$

Note that $\dfrac{1-u^2}{u}\dfrac{u^{2^p}}{1-u^{2^p}} = S_p(z)$.

Theorem 2.1. *The generating function $T_p(z)$ of trees in $\mathscr{A}$ with Horton-Strahler number $\ge p$ is for $p \ge 0$ given by*

$$T_p(z) = S_p(z)\Bigg[(1+u)^2 + (1+u+u^2)\sum_{h=1}^{p-1}\frac{1-u^{2^h}}{1-u}\prod_{j=0}^{h-1}\frac{1-u^{2^j+1}}{1-u}\Bigg]\prod_{j=0}^{p-1}\frac{1-u}{1-u^{2^j+1}}.$$

Note that this formula has been tested. Since the generating function is fully explicit one has (in principle) access to the coefficients, i. e., to the numbers of binary butterfly trees of a given number of nodes and a given Horton-Strahler number.

3. The average value of Horton-Strahler numbers in marked binary trees

By general principles, the generating function

$$\sum_{p\ge1} T_p(z)$$

is, apart from normalization, the generating function of the averages; note that it was

$$\sum_{p\ge1} S_p(z)$$

in the classical case, and the latter series is well-understood.

To understand the strategy, we need to expand the generating function around $u=1$; in some instances it helps to set $u=e^{-t}$ and expand around $t=0$. Consider

$$\Big[(1+u)^2+(1+u+u^2)\sum_{h=1}^{p-1}\frac{1-u^{2^h}}{1-u}\prod_{j=0}^{h-1}\frac{1-u^{2^j+1}}{1-u}\Big]\prod_{j=0}^{p-1}\frac{1-u}{1-u^{2^j+1}}; \tag{1}$$

the term $\dfrac{1-u^2}{u}\displaystyle\sum_{p\ge1}\frac{u^{2^p}}{1-u^{2^p}}$ will be brought in later.

The ugly term (1) is actually simpler to handle, since we are only interested in the leading term, which is a constant (in the expansion around $t=0$). For the following we might replace a factor $\frac{1-u^d}{1-u}$ by d, and instead of (1) consider

$$\lambda_p:=\Big[4+3\sum_{h=1}^{p-1}2^h\prod_{j=0}^{h-1}(2^j+1)\Big]\prod_{j=0}^{p-1}\frac{1}{2^j+1}.$$

The sequence λ_p converges to 3 exponentially fast. We consider now

$$\sum_{p\ge1}\frac{u^{2^p}}{1-u^{2^p}}\lambda_p=\sum_{p,k\ge1}\lambda_p e^{-tk2^p}.$$

The Mellin transformation will be applied, as in many related projects:

$$\mathscr{M}\sum_{p,k\ge1}\lambda_p e^{-tk2^p}=\zeta(s)\mathscr{M}\sum_{p\ge1}\lambda_p e^{-t2^p}=\zeta(s)\Gamma(s)\underbrace{\sum_{p\ge1}\lambda_p 2^{-ps}}_{\Lambda(s):=}.$$

Now

$$\Lambda(s)=\sum_{p\ge1}\lambda_p2^{-ps}=\frac{3}{2^s-1}+\sum_{p\ge1}\mathscr{O}(2^{-p})2^{-ps}.$$

The remainder has shifted singularities, and thus the dominant ones are at $\Re s=0$. The general plan is to look at the inverse Mellin transform $\zeta(s)\Gamma(s)\Lambda(s)t^{-s}$ and at its residues. Since the dominant ones are at $\Re s=0$, we can concentrate on $3\zeta(s)\Gamma(s)\frac{t^{-s}}{2^s-1}$, which is just 3 times the relevant quantity for classical binary trees.

For the final averages of the Horton-Strahler numbers, we have to divide by

$$[z^n]A(z)=\frac{3(2n)!}{(n-1)!(n+2)!}=\frac{3(2n)!}{n!(n+1)!}\Big(1+\mathscr{O}\Big(\frac1n\Big)\Big),$$

so that within the accuracy the average Horton-Strahler numbers are computed, the asymptotic formula is the same:

Theorem 3.1. *The average value of the Horton-Strahler numbers (register function) of all butterfly trees in the sense of Peca of size n is given by the asymptotic formula (with $\chi_k=\frac{2k\pi i}{\log 2}$)*

$$\begin{aligned}\log_4 n-\frac{\gamma}{2\log2}-\frac{1}{\log2}+\frac12+\log_2\pi+\frac{1}{\log2}\sum_{k\ne0}\zeta(\chi_k)\Gamma\Big(\frac{\chi_k}{2}\Big)(\chi_k-1)n^{\chi_k/2}\\=\log_4 n-\frac{\gamma}{2\log2}-\frac{1}{\log2}+\frac12+\log_2\pi+\psi(\log_4 n),\end{aligned}$$

with a tiny periodic function $\psi(x)$ of period 1.

These oscillations are usually bounded by 10^{-5}, say. See [2] for some explicit error bounds in the classical case. The remainder term in the asymptotic formula is of the form $\mathscr{O}((\log^* n)/n)$ and has never been computed explicitly, neither in the classical case nor for the butterfly trees, because of the complexity of the computations.

References

[1] N. G. de Bruijn, D. E. Knuth, and S. O. Rice, *The average height of planted plane trees*, In Graph theory and computing, pages 15–22. Academic Press, New York-London, 1972.

[2] P. Flajolet, J.-C. Raoult, and J. Vuillemin, *The number of registers required for evaluating arithmetic expressions*, Theoret. Comput. Sci. 9(1) (1979), 99–125.

[3] R. Kemp, *The average number of registers needed to evaluate a binary tree optimally*, Acta Inform. 11(4) (1978/79), 363–372.

[4] J. Peca-Medlin, *The Horton-Strahler number of butterfly trees*, 2025.

[5] H. Prodinger, *Some analytic techniques for the investigation of the asymptotic behaviour of tree parameters*, EATCS bulletin 47 (1992), 180–199.

[6] H. Prodinger, *On a problem of Yekutieli and Mandelbrot about the bifurcation ratio of binary trees*, Latin American Theoretical INformatics (Valparaíso, 1995), Volume 181, pages 181–194. 1997.

[7] H. Prodinger, *Introduction to Philippe Flajolet's work on the register function and related topics*, 2012.

[8] H. Prodinger, *k-non-crossing trees and edge statistics modulo k*, Enumer. Combin. Appl. 4:4 (2024), Article #S2R30.

[9] H. Prodinger, *Philippe Flajolet's early work in combinatorics*, Enumer. Combin. Appl. 2:1 (2022), Article #S1H1.

[10] I. Yekutieli and B. B. Mandelbrot, *Horton-Strahler ordering of random binary trees*, J. Phys. A 27(2) (1994), 285–293.