A Unified Approach to the Analysis of Horton-Strahler Parameters of Binary Tree Structures

Markus E. Nebel
Johann Wolfgang Goethe-Universität
Fachbereich Biologie und Informatik
Institut für Informatik
Frankfurt a. M.
Germany

Abstract

The Horton-Strahler number naturally arose from problems in various fields, e.g. geology, molecular biology and computer science. Consequently, detailed investigations of related parameters for different classes of binary tree structures are of interest. This paper shows one possibility of how to perform a mathematical analysis for parameters related to the Horton-Strahler number in a unified way such that only a single analysis is needed to obtain results for many different classes of trees. The method is explained by the examples of the expected Horton-Strahler number and the related r-th moments, the average number of critical nodes and the expected distance between critical nodes.

Keywords: average-case analysis, combinatorial structures, Horton-Strahler numbers, analytic combinatorics.

1 Introduction

For a binary tree T, i.e. a tree where each node has at most two descendants, the *Horton-Strahler number* $hs(T)$ is defined recursively in the following way:

$$hs(T) := \begin{cases} 0 & : \text{ if } T \text{ is either a leaf or empty} \\ hs(T.l)+1 & : \text{ if } hs(T.l) = hs(T.r) \\ \max(hs(T.l), hs(T.r)) & : \text{ otherwise} \end{cases} .$$

Here, $T.l$ (resp. $T.r$) denotes the left (resp. right) subtree of T. One example for a binary tree and its marking by the Horton-Strahler number can be found in Fig. 1. Originally this parameter was used in [16] and [28] to study the morphological structure of river networks. It is also of interest to numerous other subjects like botany or anatomy in which branching patterns appear. There are several links of the Horton-Strahler number of a binary tree to computer science. For instance, Ershov [6] has shown that the minimal number of registers needed to evaluate an arithmetic expression $\mathcal{E}$ with unary or binary operators, which is represented as a binary tree $T(\mathcal{E})$ (the syntax-tree), is given by $1 + hs(T(\mathcal{E}))$. In [7], [10] and [17] the average number of registers needed for the evaluation of syntax-trees of different types has been investigated. Furthermore, the minimum recursion-depth required for a certain traversal of a binary tree T is also given by $1 + hs(T)$ (e.g.

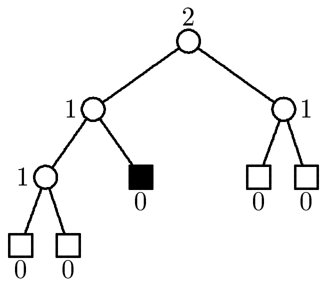


Figure 1: A binary tree marked recursively by the Horton-Strahler number.

see [14]). Meir, Moon and Pounder [21] investigated the Horton-Strahler number of channel networks with a fixed number of inputs. Prodinger has investigated parameters like the average number of critical nodes [26] or the register path length of binary trees where only critical nodes and not all nodes contribute to the path length [27]. Here, a node is called *critical* if its two subtrees possess the same value of hs, i.e. if it is responsible for a growth of the Horton-Strahler number. Recently, the author determined the average Horton-Strahler number of a class of trees (called $\mathcal{C}$-tries) which model the trie data structure in a combinatorial way [23]; the Horton-Strahler number of tries in the Bernoulli model has been investigated in [5]. The combinatorics of the Horton-Strahler analysis has been used in computer graphics for the creation of faithful synthetic images of natural trees (see [31]) and for information visualization [15]. The Horton-Strahler number also appears in molecular biology in connection with some theoretical considerations about secondary structures of single-stranded nucleic acids [30]. With respect to this application we usually speak about the *order* of a secondary structure instead of its Horton-Strahler number. In [25] it is shown how the generating functions which will be derived here can be used to asymptotically enumerate the number of secondary structures of order p built from n bases. This solves a problem established by Waterman [29] in 1978. A survey on the application of Horton-Strahler parameters in different scientific areas has been presented in [32].

All the previously cited studies presenting a mathematical analysis of a Horton-Strahler parameter (e.g. average value, variance, etc.) have in common that they choose a fixed family of trees and a fixed parameter and then perform a dedicated analysis for that setting. In the sequel a unified approach will be presented allowing the investigation of a Horton-Strahler parameter for various families of trees by just one analysis. The key for this approach is using multivariate generating functions in such a way that appropriate substitutions for the variables can be finally used to switch over to the different families of trees. The amount of work that will have to be done after this substitution depends on the number of parameters that are required in the final result. If the result possesses only a single parameter (e.g. the total number of nodes of the tree), then no additional work is necessary, if multiple parameters appear (e.g. the number of internal nodes and the number of leaves), then an additional coefficient of our generating functions has to be determined. In this way we will derive results for the following families of binary trees:

- Extended binary trees;
- Motzkin trees;
- unary/binary trees with c_1 different types of unary nodes and c_2 different types of binary nodes;
- $\mathcal{C}$-tries.

We will consider two notions of size, namely trees with n nodes and trees with n nodes and ℓ (non-empty) leaves. The results are derived assuming that all trees of the same size are equally likely.

The approach presented makes it possible to perform unified computations for different parameters like the expected Horton-Strahler number (and higher moments), the average number of critical nodes and the expected distance between critical nodes. In the sequel we will use $\bigcirc$ to represent an internal node of a tree; $\square$ will be used for the representation of a leaf. The distinction between different kinds of leaves (e.g. leaves which really exist and leaves which represent an empty position of the tree) will be done by coloring. For example, the black leaf in Fig. 1 could be considered as a NIL-pointer. The notation $[x_1^{n_1} \cdots x_k^{n_k}]f(x_1, \ldots, x_k)$ is used to represent the coefficient at $x_1^{n_1} \cdots x_k^{n_k}$ in the expansion of $f(x_1, \ldots, x_k)$ at $(x_1, \ldots, x_k) = (0, \ldots, 0)$, $k \geq 1$.

2 A Unified Analysis of Horton-Strahler Parameters

In this section we will present a unified approach to the analysis of Horton-Strahler parameters of binary tree structures. The method will be explained using the expected Horton-Strahler number and the related higher moments, the average number of critical nodes and the average distance between critical nodes as examples. We will proceed in two steps. First, the idea of our approach will be discussed, then the mathematical details will be given.

For any extended binary tree structure we can distinguish three types of internal nodes. First, there are nodes where both successors are leaves (type 0); second, there are internal nodes where one successor is a leaf but the other one is not (type 1) and third, there are internal nodes where none of the successors are leaves (type 2). Note that for extended trees an internal node always possesses two non-empty successors. However, it will also be possible to consider families of trees in which an internal node may have empty successors by appropriate substitutions. We can generate different families of binary tree structures by weighting the internal nodes of each type in different ways. For example, if the nodes of type 0 and 2 are weighted by one and those of type 1 by two, then we find the class of $\mathcal{C}$-tries introduced in [22]. This can be seen very easily by using symbolic equations as proposed in [8]. We start with the binary tree structure $\mathcal{U}$ characterized by the equation

$$\mathcal{U} = \underbrace{\overset{\bigcirc}{\overset{/\ \backslash}{\square\ \ \square}}}_{\text{type } 0} + \underbrace{\overset{\bigcirc}{\overset{/\ \backslash}{\square\ \ \mathcal{U}}} + \overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{U}\ \ \square}}}_{\text{type } 1} + \underbrace{\overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{U}\ \ \mathcal{U}}}}_{\text{type } 2}. \tag{1}$$

By assigning the weights mentioned above, equation (1) reads now

$$\mathcal{U}' = \overset{\bigcirc}{\overset{/\ \backslash}{\square\ \ \square}} + 2 \cdot \overset{\bigcirc}{\overset{/\ \backslash}{\square\ \ \mathcal{U}'}} + 2 \cdot \overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{U}'\ \ \square}} + \overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{U}'\ \ \mathcal{U}'}}.$$

We may now assume that the factor 2 for a node belonging to type 1 results from two different possibilites for the leaf attached to it. Then we end up with

$$\mathcal{C} = \overset{\bigcirc}{\overset{/\ \backslash}{\square\ \ \square}} + \overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{C}\ \ \{\square, \blacksquare\}}} + \overset{\bigcirc}{\overset{/\ \backslash}{\{\square, \blacksquare\}\ \ \mathcal{C}}} + \overset{\bigcirc}{\overset{/\ \backslash}{\mathcal{C}\ \ \mathcal{C}}}, \tag{2}$$

where $\{\square, \blacksquare\}$ is used to represent the two different possibilities for that leaf. But this is exactly the symbolic equation for the familiy of $\mathcal{C}$-tries, where $\blacksquare$ is used to

represent a NIL pointer and $\square$ pictures a leaf which stores a key.
In some sense this concept is related to the notion of *simply generated families of trees* introduced in [20] and ideas used in [10]. The mathematical treatment using generating functions is almost straightforward. We introduce a variable for each type of node. Let v (resp. u, x) mark an internal node of type 0 (resp. 1, 2) and translate the symbolic equation (1) into the corresponding equation for the enumerator generating function $T(x,u,v)$. We find

$$T(x,u,v) = v + 2uT(x,u,v) + xT(x,u,v)^2$$

and thus

$$T(x,u,v) = \frac{1-2u-\sqrt{1-4u+4u^2-4xv}}{2x}.$$

According to our previous considerations, the enumerator generating function for $\mathcal{C}$-tries, where each internal node is marked by z, is given by $T(z,2z,z)$ and is therefore equal to $\frac{1-4z-\sqrt{1-8z+12z^2}}{2z}$.
In order to derive results for Horton-Strahler parameters we need a representation of the generating function $R_p(x,u,v)$ which counts those trees that have a Horton-Strahler number equal to p. From the definition of $hs(T)$ we can conclude that for $p \geq 2$

$$R_p(x,u,v) = 2uR_p(x,u,v) + xR_{p-1}^2(x,u,v) + 2xR_p(x,u,v) \sum_{1\leq j<p} R_j(x,u,v)$$

holds. Furthermore, it is obvious that $R_1(x,u,v) = \frac{v}{1-2u}$. As in the specific case of extended binary trees [7] or combinatorial tries [23] it is possible to solve this recursion by a trigonometric change of variables. In that way it is possible to prove that

$$R_p(x,u,v) = \frac{-1}{U_{2^p-1}(\xi)} \frac{v}{\sqrt{xv}},\ p \geq 1, \tag{3}$$

where $U_n(z)$ is the n-th Chebyshev polynomial of the second kind [1, 22.2.5] and $\xi := \frac{2u-1}{2\sqrt{xv}}$. By the application of the closed form representation of $U_n(z)$ given in [18, B74] and some obvious simplifications we finally find a much more appropriate form, namely

$$R_p(x,u,v) = \frac{v}{\sqrt{xv}} \frac{(1-\omega)\omega^{2^{p-1}}}{\sqrt{\omega}(1-\omega^{2^p})},\ p \geq 1, \tag{4}$$

where $\omega := \frac{1-\varepsilon}{1+\varepsilon}$, $\varepsilon := \sqrt{1-4\frac{xv}{(1-2u)^2}}$. Furthermore we will need the generating function $S_p(x,u,v)$ of trees with a Horton-Strahler number of at least p. It turns out that

$$\begin{aligned} S_p(x,u,v) &= -\frac{1-2u+\sqrt{1-4u+4u^2-4xv}}{2x} - \frac{v}{\sqrt{xv}} \frac{U_{2^{p-1}}(\xi)}{U_{2^{p-1}-1}(\xi)} \\ &= \frac{v}{\sqrt{xv}} \frac{1-\omega}{\sqrt{\omega}} \frac{\omega^{2^{p-1}}}{1-\omega^{2^{p-1}}} \end{aligned} \tag{5}$$

holds. Based on these representations for $R_p(x,u,v)$ and $S_p(x,u,v)$ it becomes possible to investigate different Horton-Strahler parameters of binary trees in a unified way.

2.1 The Expected Horton Strahler Number

In order to determine the expected Horton-Strahler number of a binary tree we have to consider the generating function

$$M(x,u,v) := \sum_{p\geq 1} pR_p(x,u,v).$$

By definition of S_p we have $\sum_{p\geq 1} pR_p(x,u,v) = \sum_{p\geq 1} S_p(x,u,v)$ and thus by (5)

$$M(x,u,v) = \frac{v}{\sqrt{xv}} \frac{1-\omega}{\sqrt{\omega}} \sum_{p\geq 0} \frac{\omega^{2^p}}{1-\omega^{2^p}} = \frac{v}{\sqrt{xv}} \frac{1-\omega}{\sqrt{\omega}} \underbrace{\sum_{\substack{p\geq 0\\ \lambda\geq 0}} \omega^{(\lambda+1)2^p}}_{=:\varsigma(\omega)} \tag{6}$$

holds. We will first consider the series $\varsigma(\omega)$ appearing in (6). The analysis of harmonic summations like $\varsigma(\omega)$ is performed by means of the Mellin transform which is by now a fairly well understood methodology in analytic combinatorics and analysis of algorithms (see for instance the excellent survey by Flajolet et al. [13]) going back to the seminal paper of De Bruijn et al. [3]. We determine the Mellin transform of $\varsigma(e^{-t})$ which proves to be given in closed form by

$$\eta(s) := \Gamma(s)\zeta(s)\frac{1}{1-2^{-s}},\ \Re(s) > 1,$$

where $\Gamma(s)$ is the complete gamma function and $\zeta(s)$ is the Riemann zeta function. Then, according to the methodology, an expansion of $\varsigma(e^{-t})$ around $t=0$ is given by the sum of residues of $t^{-s}\eta(s)$. We have a pole for $s=1$ because of the zeta function, poles at $s=-k$, $k\in\mathbb{N}_0$, due to the gamma function and poles at $s=\chi_m := \frac{2\pi i m}{\ln(2)}$, $m\in\mathbb{Z}$, introduced by $\frac{1}{1-2^{-s}}$. It is well-known that the poles at χ_m for $m\neq 0$ will only introduce an oscillation of very small amplitude in the final result. For the sake of simplicity we will not consider them, although it is not difficult to extend the approach presented here to deal with them. The poles at $s=-k$, $k\geq 1$, are responsible for lower order terms only, therefore we will not consider them either. The sum of the remaining residues is given by

$$2t^{-1} + \frac{1}{2}\log_2(t) + \frac{2\gamma - 2\ln(\pi) - 3\ln(2)}{4\ln(2)}.$$

At this point we have to make a fundamental decision, namely to choose the nodes (the types of nodes) which we want to contribute to the size in our final result. Here we will assume that each internal node contributes and thus we set x to xz, u to uz and v to vz. Furthermore, we assume that x, u and v are fixed real *parameters* greater than 0. Then we have the following trivial bounds for $[z^n]M(xz,uz,vz)$:

$$[z^n]T(xz,uz,vz) \leq [z^n]M(xz,uz,vz) \leq \log_2(n+1)[z^n]T(xz,uz,vz).$$

Since the Cauchy-Hadarmard theorem tells us that the minorant and the majorant have the same radius of convergence $r = |\frac{1}{2u+2\sqrt{xv}}|$, we conclude that r is the radius of convergence of $M(xz,uz,vz)$ as a function in z. Furthermore, the theorem of Pringsheim implies that $M(xz,uz,vz)$ has a singularity at $z=\frac{1}{2u+2\sqrt{xv}}$. The $\mathcal{O}$-tranfer method [9] approximates the coefficient $[z^n]$ of a generating function with dominant singularity z_d by means of Cauchy's formula together with the contour $\mathcal{C}$ depicted in Fig. 2. Therefore, in order to apply the method we need an analytic continuation of $M(xz,uz,vz)$ outside its radius of convergence. It is obvious that this continuation exists since, with respect to the ω-plane, $M(xz,uz,vz)$ has the radius of convergence 1 and the contour $\mathcal{C}_\omega$ in the ω-plane which is equivalent to the contour $\mathcal{C}$ in the z-plane lies completely inside the unit circle. Thus we can use an expansion of our generating function at $z=z_d := \frac{1}{2u+2\sqrt{xv}}$ to derive an asymptotic for the coefficient $[z^n]$. To get this expansion we have to resubstitute t within the sum of residue given above. Since $t=-\ln((1-\varepsilon)/(1+\varepsilon))$ and $\varepsilon=0$ for $z=z_d$ we expand $-\ln((1-\varepsilon)/(1+\varepsilon))$ around $\varepsilon=0$ yielding

$$t \sim 2\sqrt{2+2\frac{u}{\sqrt{xv}}}\sqrt{1-2z(u+\sqrt{xv})}. \tag{7}$$

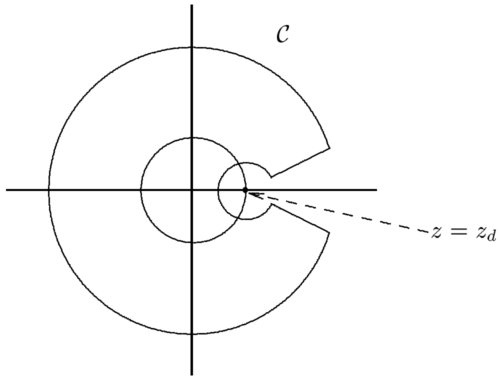


Figure 2: The contour $\mathcal{C}$ used by the $\mathcal{O}$-transfer method.

Furthermore, we have to take care of the factor $\frac{v}{\sqrt{xv}} \frac{1-\omega}{\sqrt{\omega}}$ which expands to

$$\frac{2v\sqrt{2u+2\sqrt{xv}}}{(xv)^{3/4}} \sqrt{1-2z(\sqrt{xv}+u)}. \tag{8}$$

Thus the most significant terms of the expansion of $M(xz, uz, vz)$ at $z = z_d$ are given by

$$\frac{v}{2(xv)^{3/4}} \sqrt{2u+2\sqrt{xv}} \sqrt{1-2z(\sqrt{xv}+u)} \log_2(1-2z(\sqrt{xv}+u))$$

$$+ \frac{v}{4(xv)^{3/4}\ln(2)} \sqrt{2u+2\sqrt{xv}} \left(\ln\left(\frac{(\sqrt{xv}+u)^2}{\pi^4 xv} \right) + 4\gamma \right) \sqrt{1-2z(\sqrt{xv}+u)}.$$

By means of the $\mathcal{O}$-transfer this expansion can be translated into an asymptotic for the coefficient $[z^n]M(xz, uz, vz)$. We find

Theorem 1 *Let $M(x,u,v) = \sum_{p\geq 1} pR_p(x,u,v)$ for $R_p(x,u,v)$ the ordinary generating function of extended binary trees with Horton-Strahler number p and each internal node of type 0 (resp. type 1, type 2) marked by v (resp. u, x). Then*

$$[z^n]M(xz, uz, vz) \sim$$
$$\frac{1}{\sqrt{\pi n^3}} (2u+2\sqrt{xv})^n \frac{v\sqrt{2u+2\sqrt{xv}}}{4(xv)^{3/4}\ln(2)} \left(\ln(n) - \ln\left(\frac{\sqrt{xv}+u}{4\pi^2\sqrt{xv}} \right) - \gamma - 2 \right),$$

for fixed real x, u, $v > 0$ and $n \to \infty$. □

We will use this asymptotic formula to derive results for explict families of binary trees in the next section.
Before we will consider the critical nodes we have a quick look at higher moments of the expected Horton-Strahler number. For that purpose we regard

$$M^{(r)}(x,u,v) := \sum_{p\geq 1} p^r R_p(x,u,v) = \frac{v}{\sqrt{xv}} \frac{1-\omega}{\sqrt{\omega}} \underbrace{\sum_{\substack{p\geq 1\\ j\geq 0}} p^r \omega^{2^{p-1}(1+2j)}}_{=:\varsigma^{(r)}(\omega)}.$$

Again, we use the Mellin summation technique to find an expansion of the harmonic sum around the dominant singularity. The Mellin transform of $\varsigma^{(r)}(e^{-t})$ is given by

$$\mathcal{M}^{(r)}(s) := \frac{\Gamma(s) 2^s A_r(2^{-s}) \zeta(s)}{(1-2^{-s})^r},$$

where $A_n(x)$ denotes the n-th Eulerian polynomial for which [4, p. 245]

$$\sum_{l\geq 0} l^n u^l = \frac{A_n(u)}{(1-u)^{n+1}}$$

holds. There is a pole at $s=0$ of order $r+1$ which is responsible for the most significant contribution to the asymptotic of the coefficient. Therefore we are interested in the residue of $t^{-s}\mathcal{M}^{(r)}(s)$ at $s=0$. Since for $s \to 0$

$$\begin{aligned}
A_r(2^{-s}) &= r!, \text{ see [19, 5.1.3(4)]},\\
\zeta(s) &= -\frac{1}{2},\\
\Gamma(s) &= s^{-1} - \gamma + \mathcal{O}(s),\\
\frac{1}{(1-2^{-s})^r} &= \frac{1}{\ln^r(2)} s^{-r} + \mathcal{O}(s^{-r+1}), \text{ and}\\
t^{-s} &= \sum_{i\geq 0} \ln^i(t) \frac{(-1)^i}{i!} s^i,
\end{aligned}$$

we can conclude that the residue of $t^{-s}\mathcal{M}^{(r)}(s)$ at $s=0$ is given by

$$(-1)^{r+1}\frac{\ln^r(t)}{2\ln^r(2)} + \mathcal{O}(\ln^{r-1}(t)). \tag{9}$$

By resubstituting t by (7) and multiplying the resulting expression by the expansion (8), we get the most significant term of the expansion of $M^{(r)}(xz, uz, vz)$ around its dominant singularity z_d

$$-2^{-r}\frac{v}{(xv)^{3/4}}\sqrt{2u+2\sqrt{xv}}\sqrt{1-2z(\sqrt{xv}+u)}\log_2^r\left(\frac{1}{1-2z(u+\sqrt{xv})}\right).$$

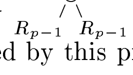
Finally the application of the $\mathcal{O}$-transfer method yields

Theorem 2 *Let $M^{(r)}(x,u,v) = \sum_{p\geq 1} p^r R_p(x,u,v)$ for $R_p(x,u,v)$ the ordinary generating function of extended binary trees with Horton-Strahler number p and each internal node of type 0 (resp. type 1, type 2) marked by v (resp. u, x). Then*

$$[z^n]M^{(r)}(xz,uz,vz) \sim \frac{2^{-r-1}(2u+2\sqrt{xv})^n}{\sqrt{\pi n^3}}\log_2^r(n)\frac{v}{(xv)^{3/4}}\sqrt{2u+2\sqrt{xv}},$$

for fixed real x, u, $v > 0$ and $n \to \infty$. □

2.2 The Average Number of Critical Nodes

In order to get an expression for the average number of critical nodes we have to derive from $R_p(x,u,v)$ a representation of $K(x,u,v) := \sum_{p\geq 1} K_p(x,u,v)$, where $K_p(x,u,v)$ denotes the ordinary generating function for the total number of critical nodes with a Horton-Strahler number p. A formula for $K_p(x,u,v)$ can be derived by following the idea presented in Fig. 3. If we have an arbitrary tree with a subtree $\triangle$ then we can introduce an additional critical node with mark p by substituting a leaf of $\triangle$ by 

. Each tree T with k different critical nodes with mark p will be generated by this procedure in k different ways. Thus the resulting generating

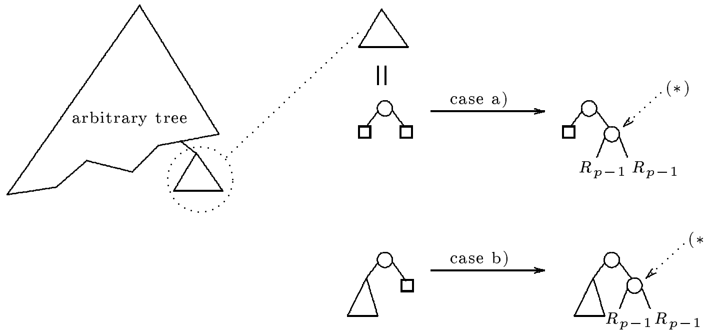


Figure 3: How to construct an arbitrary tree with a critical node $(*)$ with mark p in a specific position.

function will give us the total number of critical nodes with mark p. The procedure just described translates into generating functions in the following way:

$$\begin{aligned} K_p(x,u,v) &= \underbrace{\frac{\partial}{\partial a} T(x,u,va^2)\,|_{a=1}\, \frac{ux}{v} R_{p-1}^2(x,u,v)}_{\text{case a)}} \\ &\quad \underbrace{+\frac{\partial}{\partial b} T(x,ub,v)\,|_{b=1}\, \frac{x^2}{u} R_{p-1}^2(x,u,v)}_{\text{case b)}} + xR_{p-1}^2(x,u,v) \\ &= \frac{x}{\sqrt{1-4u+4u^2-4xv}} R_{p-1}^2(x,u,v),\ p \geq 1. \end{aligned}$$

The term $xR_{p-1}^2(x,u,v)$ is needed because by construction $T(x,u,v)$ does not count the empty tree $\square$. Please note that at the first sight this derivation seems to be valid only for $p \geq 2$ since for $p = 1$ the internal node of $R_{p-1}\ R_{p-1}$ is of type 0 and thus should be marked by v instead of x. Fortunately, for both representations (3) and (4) of $R_p(x,u,v)$ we find that $xR_0^2(x,u,v) = v$ holds. Thus there is no need for a separated discussion on the case $p = 1$. Therefore, by (4), we find the following representation of the generating function in question:

$$K(x,u,v) = \underbrace{\frac{v}{\sqrt{1-4u+4u^2-4xv}} \frac{(1-\omega)^2}{\omega}}_{=:\phi} \underbrace{\sum_{\substack{p\geq 0\\ \lambda \geq 1}} \lambda \omega^{\lambda 2^p}}_{=:\sigma(\omega)}. \tag{10}$$

We will first consider the summation $\sigma(\omega)$ of (10). We determine the Mellin transform of $\sigma(e^{-t})$ which proves to be given in closed form by

$$\eta(s) := \Gamma(s)\zeta(s-1)\frac{1}{1-2^{-s}},\ \Re(s) > 2.$$

The transform $\eta(s)$ has a pole for $s = 2$ because of the zeta function, poles at $s = -k$, $k \in \mathbb{N}_0$, due to the gamma function and poles at $s = \chi_m := \frac{2\pi i m}{\ln(2)}$, $m \in \mathbb{Z}$, introduced by $\frac{1}{1-2^{-s}}$. Again, the poles at χ_m for $m \neq 0$ will only introduce an oscillation of very small amplitude in the final result. The poles at $s = -k$, $k \geq 1$, are responsible for lower order terms only, therefore we will not consider them. The sum of the remaining residues is given by

$$\frac{4}{3}t^{-2} + \frac{1}{12}\log_2(t).$$

In accordance with the analysis of the expected Horton-Strahler number we will assume that each internal node contributes and thus we set x to xz, u to uz and v to vz. Again we assume that x, u and v are fixed real parameters > 0. By the same reasoning as before

$$[z^n]T(xz,uz,vz) \le [z^n]K(xz,uz,vz) \le n[z^n]T(xz,uz,vz),$$

implies that $K(xz,uz,vz)$ possesses the dominant singularity $z_d = \frac{1}{2u+2\sqrt{xv}}$. We use (7) in order to resubstitute t which yields

$$\frac{1}{3}\frac{1}{2+2\frac{u}{\sqrt{xv}}}(1-2z(u+\sqrt{xv}))^{-1} + \frac{1}{24}\log_2(1-2z(u+\sqrt{xv}))$$

for the expansion of $\sigma(\omega)$ around z_d. Now we can consider the factor ϕ in (10) which has to be expanded around z_d also. We find the expansion

$$\frac{2}{x}\sqrt{2xv+2u\sqrt{xv}}\sqrt{1-2z(u+\sqrt{xv})}.$$

Finally we recombine our results according to (10) in order to get the expansion of $K(xz,uz,vz)$ around the dominant singularity. This expansion can be translated into an asymptotic for $[z^n]K(xz,uz,vz)$ by means of the $\mathcal{O}$-transfer method which proves the subsequent theorem.

Theorem 3 *Let $K(x,u,v)$ be the ordinary generating function of the total number of critical nodes in all extended binary trees with each internal node of type 0 (resp. type 1, type 2) marked by v (resp. u, x). Then*

$$\begin{aligned}[z^n]K(xz,uz,vz) \quad \sim \quad & \frac{1}{3\sqrt{\pi n}}\sqrt{\frac{v}{x}}\frac{\sqrt{2xv+2u\sqrt{xv}}}{\sqrt{xv}+u}\left(2u+2\sqrt{xv}\right)^n \\ & + \frac{\sqrt{2xv+2u\sqrt{xv}}}{24x\sqrt{\pi n^3}}\left(2u+2\sqrt{xv}\right)^n\log_2(n),\end{aligned}$$

for fixed real x, u, $v > 0$ and $n \to \infty$. □

We can also consider the critical nodes with a fixed Horton-Strahler number p. For that purpose we regard $K_p(xz,uz,vz)$. Using the representation (3) of R_p together with [1, 22.16.5]

$$U_n\left(\cos\left(\frac{m\pi}{n+1}\right)\right) = 0,\ 1 \le m \le n,$$

we find that the dominant singularity of $K_p(xz,uz,vz)$ with respect to z is not determined by the poles introduced by the Chebyshev polynomial but by the branching-point $z = z_d = \frac{1}{2u+2\sqrt{xv}}$ of the factor $(1-4uz+4u^2z^2-4xvz^2)^{-1/2}$. The expansion of $K_p(xz,uz,vz)$ at z_d is given by

$$2^{1/2-2p}\frac{v^{3/4}}{\sqrt{u+\sqrt{xv}}x^{1/4}}(1-2z(u+\sqrt{xv}))^{-1/2}$$

such that we find the following theorem.

Theorem 4 *Let $K_p(x,u,v)$ be the ordinary generating function of the total number of critical nodes with Horton-Strahler number p in all extended binary trees with each internal node of type 0 (resp. type 1, type 2) marked by v (resp. u, x). Then*

$$[z^n]K_p(xz,uz,vz) \sim 2^{1/2-2p}\frac{v^{3/4}}{\sqrt{\pi n}\sqrt{u+\sqrt{xv}}x^{1/4}}(2u+2\sqrt{xv})^n,$$

for fixed real x, u, $v > 0$ and $n \to \infty$. □

An asymptotic of increased precision would result from additional terms of the expansion of $K_p(xz,uz,vz)$ at z_d which can easily be determined if required.

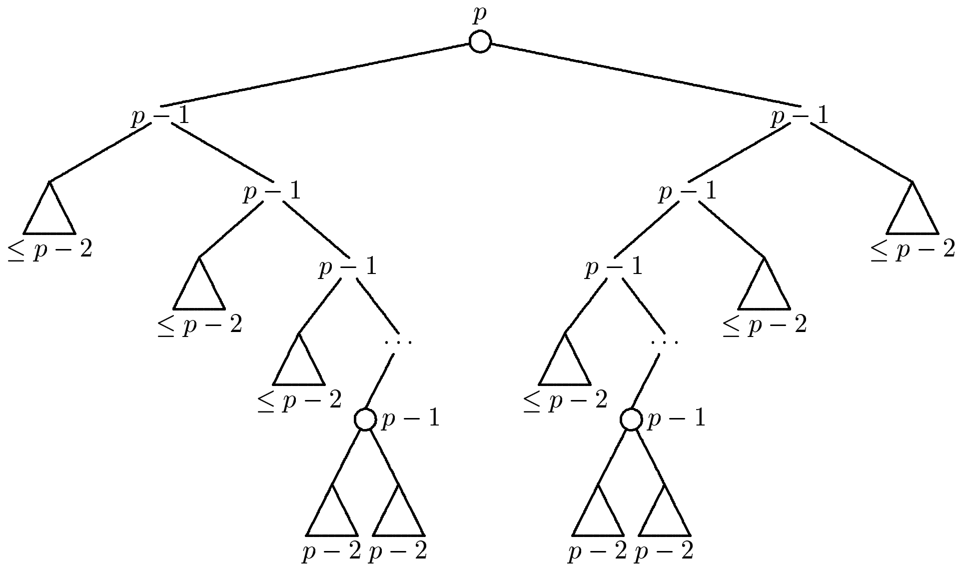


Figure 4: A tree with a critical root marked by p. Each critical node is represented by $\bigcirc$, $\triangle$ is used to represent an entire subtree with a root-mark as given below it.

2.3 The Length of Critical Paths

Let a path from a critical node with Horton-Strahler number p to one of its critical successors with Horton-Strahler number $p-1$ be called *critical path of order p*. The next parameter that we want to study is the total length of all critical paths of order p. For that purpose we derive a generating function which marks each corresponding edge by variable w. We first restrict ourselves to the case of a tree τ with critical root marked by p as shown in Fig. 4. For that tree exactly the edges on the paths from the two critical nodes $\bigcirc$ and mark $p-1$ to the root of the tree contribute to the length and thus have to be marked by w. In the general case, each subtree of a critical root marked by p consists of a list of zero or more non critical nodes and a succeeding critical node all marked by $p-1$. Thus, for $P_p(x,u,v) := T(x,u,v) - S_p(x,u,v)$ and $p \geq 2$ all possible subtrees are enumerated by

$$\tau_{p-1} := R^2_{p-2}(x,u,v)\frac{x}{1-2wx(\frac{u}{x}+P_{p-1}(x,u,v))}.$$

The entire tree τ then is given by $xw^2\tau^2_{p-1}$. Since critical nodes with a Horton-Strahler number p may not only appear as root of a tree, τ must be embedded into a larger tree as a substructure. This can be done by the same ideas used for the derivation of $K(x,u,v)$. We find

$$\frac{xw^2}{\sqrt{1-4u+4u^2-4xv}}\left(R^2_{p-2}(x,u,v)\frac{x}{1-2wx(\frac{u}{x}+P_{p-1}(x,u,v))}\right)^2$$

for the resulting generating function. We take the first partial derivative with respect to w and afterwards set $w := 1$ to find a representation of the generating function $KD_p(x,u,v)$ enumerating all edges on critical paths of order p. Appropriate simplifications lead to

$$KD_p(x,u,v) \quad = \quad \frac{2xR^2_{p-1}(x,u,v)}{\sqrt{1-4u+4u^2-4xv}}\frac{1}{1-2u-2xP_{p-1}(x,u,v)}$$

$$= \frac{2R_{p-1}^3(x,u,v)}{R_{p-2}^2(x,u,v)} \frac{1}{\sqrt{1-4u+4u^2-4xv}}.$$

For both equations we used the identity $R_p(x,u,v) = \tau_p|_{w=1}$. By the same reasoning as used for K_p we find that $z = z_d = \frac{1}{2u+2\sqrt{xv}}$ is the dominant singularity of $KD_p(xz,uz,vz)$ implied by the branching-point of $\sqrt{1-4uz+4u^2z^2-4xvz^2}$. The leading term of the expansion of $KD_p(xz,uz,vz)$ at z_d proves to be given by

$$2^{-p-1/2} \frac{v^{1/4}\sqrt{u+\sqrt{xv}}}{x^{3/4}} (1-2z(u+\sqrt{xv}))^{-1/2}$$

such that the $\mathcal{O}$-transfer method leads to

Theorem 5 *Let $KD_p(x,u,v)$ be the ordinary generating function counting the total distance (measured in the number of edges) between all critical nodes with Horton-Strahler number p and their critical successors with Horton-Strahler number $p-1$ in all extended binary trees with each internal node of type 0 (resp. type 1, type 2) marked by v (resp. u, x). Then*

$$[z^n]KD_p(xz,uz,vz) \sim 2^{-p-1/2} \frac{v^{1/4}\sqrt{u+\sqrt{xv}}}{\sqrt{\pi n}x^{3/4}} (2u+2\sqrt{xv})^n,$$

for fixed real x, u, $v > 0$ and $n \to \infty$. □

3 Results for Specific Families of Trees

In this section we will use our general results obtained in the previous section in order to determine the average Horton-Strahler number and the related higher moments, the average number of critical nodes with respect to the Horton-Strahler number, and the expected distance between critical nodes, for certain families of binary trees. In this way it will be possible to prove new results, for many old result our technique will lead to a new proof. Since we will be assuming the uniform distribution in each family of trees, the expected Horton-Strahler number h_n of trees of size n is given by

$$h_n = \frac{[z^n]M(z,z,z)}{[z^n]T(z,z,z)}.$$

Similarly, the expected Horton-Strahler number $h_{n,\ell}$ of trees of size n with ℓ leaves would be given by

$$h_{n,\ell} = \frac{[z^n u^\ell]M(z,zu,zu^2)}{[z^n u^\ell]T(z,zu,zu^2)}.$$

In this case we must pay attention since now u cannot be assumed to be a kind of fixed parameter. Obviously, formulæ of the same pattern exist for the other parameters considered in the second section. However, we will have to make minor adjustments to this general procedure for Motzkin trees, unary-binary trees and $\mathcal{C}$-tries as we have derived our generating functions only for families of binary trees with non-empty leaves.

As mentioned above, the singularities χ_m and the resulting residues would imply an oscillation of very small amplitude. In the case of $[z^n]M(z,z,z)$ this oscillation is of order $\mathcal{O}(4^n n^{-3/2})$ and thus of order $\mathcal{O}(1)$ for the expectation. The same remark also holds for the other families of trees that will be considered in this paper. The presence of the oscillation will be pointed out by a term $\Delta(n)$ within the corollaries without giving any representation for Δ (even though our method is general enough to compute one). Therefore the constant term of the expected Horton-Strahler number given in the succeeding corollaries is not precise in a rigorous sense. Note that Δ is not necessarily the same function for all corollaries. Note also that the same remarks would be true for the constant term of the expected number of critical nodes, if we would use our method to determine it.

Before we can derive results for specific families of trees we need an asymptotical representation of the number of trees of size n. This number can be determined by the same methods starting at $T(xz, uz, vz)$. The expansion of $T(xz, uz, vz)$ around its dominant singularity z_d translates into

$$[z^n]T(xz, uz, vz) \sim \frac{\sqrt{2xv + 2u\sqrt{xv}}}{2x\sqrt{\pi n^3}} \left(2u + 2\sqrt{xv}\right)^n.$$

Now we are ready to consider specific families of binary trees.

3.1 Extended Binary Trees

In order to get results for the familiy of extended binary trees with n internal nodes we have to set $x := 1$, $u := 1$ and $v := 1$ in our asymptotics of the Theorems 1, 2, 3, 4 and 5 and divide the corresponding expression by $[z^n]T(z, z, z)$. In this way we find:

Corollary 1 *On the assumption that all extended binary trees of the same size (number of internal nodes) are equally likely, the average Horton-Strahler number of a tree of size n is asymptotically given by [7], [17]*

$$\log_4\left(2\pi^2 n\right) - \frac{2+\gamma}{2\ln(2)} + \Delta(n),\ n \to \infty.$$

The corresponding r-th moments are given by

$$2^{-r}\log_2^r(n),\ n \to \infty.$$

The average number of critical nodes is asymptotically given by [26]

$$\frac{1}{3}n + \frac{1}{12}\log_2(n),\ n \to \infty.$$

On the average, there are

$$4^{-p}n$$

critical nodes with fixed Horton-Strahler number p in an extended binary tree of size n, $n \to \infty$, which have an expected distance of

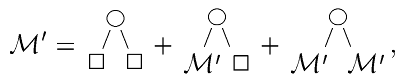
$$2^p$$

to their critical successors. □

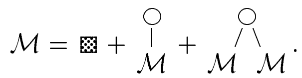
Note, that all terms depending on x, u and v in the second-order term of the total number of critical nodes vanish in the expression giving the expected value. This implies that the average number of critical nodes possesses the same term of second order for all families of binary trees that can be investigated by our approach.

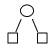
3.2 Motzkin Trees

In this section we will consider Motzkin trees, i.e. ordered trees for which an internal node may have one or two successors. We can obtain results for that familiy of trees by setting $x := 1$, $u := \frac{1}{2}$ and $v := 1$. By this substitution the structure of the trees counted by our generating functions becomes



while the Motzkin trees are described by



The non-empty leaves □ in the trees enumerated by the generating functions of section 2 introduce two problems we have to cope with: First, the Horton-Strahler number of a Motzkin tree and all its nodes is shifted by one, since the presence of the leaves □ implies that the leaves of the corresponding Motzkin tree are labeled by 1. Second, the number of critical nodes is overestimated by the number of leaves of the Motzkin trees, since each leaf ▩ in a Motzkin tree is represented by  in our computation and thus is counted as critical. Fig. 5 shows a Motzkin tree together with the extended binary tree which is considered to be equivalent in our generating functions (after setting $x := 1$, $u := \frac{1}{2}$ and $v := 1$). However, we can work around

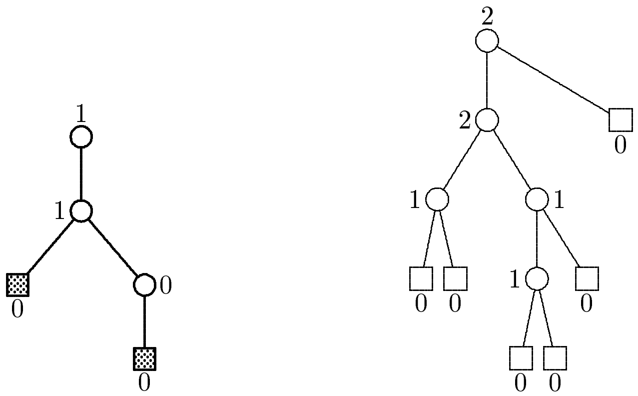


Figure 5: A Motzkin tree (left image) together with an extended binary tree (right image) which are assumed to be equivalent by the substitution $x = 1$, $u = \frac{1}{2}$ and $v = 1$.

these problems very easily. For the expected Horton-Strahler number we just have to subtract 1 from our result, since the Horton-Strahler number of every tree is overestimated by 1. For the r-th moments we need to consider $\sum_{p\geq 1} p^r R_{p+1}(x,u,v)$ instead of $\sum_{p\geq 1} p^r R_p(x,u,v)$. Obviously, the Mellin-transform of this new harmonic sum differs from the one considered in Theorem 2 just by the factor 2^{-s} and thus, since the asymptotic of Theorem 2 results from the pole at $s = 0$, our leading term remains unaffected. For the total number of critical nodes we consider $\overline{K}(x,u,v) := \sum_{p\geq 2} K_p(x,u,v)$ instead of $K(x,u,v) = \sum_{p\geq 1} K_p(x,u,v)$. This modification implies just a small change for our computations. The Mellin transform of the resulting sum $\overline{\sigma}(e^{-t})$ is $\Gamma(s)\zeta(s-1)\frac{2^{-s}}{1-2^{-s}}$ with the sum of residues $\frac{1}{3}t^{-2} + \frac{1}{12}\log_2(t)$. Thus we only have to multiply the leading term of our result by $\frac{1}{4}$ in order to get the correct result for the Motzkin trees. For K_p and KD_p we just have to set p to $p+1$ to take care of the shift. We find in this way:

Corollary 2 *On the assumption that all Motzkin trees of the same size (total number of nodes) are equally likely, the average Horton-Strahler number of a Motzkin* *tree of size n is asymptotically given by [10]*

$$\log_4\left(\frac{2}{3}\pi^2 n\right) - \frac{2+\gamma}{2\ln(2)} + \Delta(n),\ n \to \infty.$$

The corresponding r-th moments are given by

$$2^{-r}\log_2^r(n),\ n \to \infty.$$

The average number of critical nodes of a Motzkin tree of size n is asymptotically given by

$$\frac{1}{9}n + \frac{1}{12}\log_2(n),\ n \to \infty.$$

On the average, there are

$$\frac{1}{3}4^{-p}n$$

critical nodes with fixed Horton-Strahler number p in a Motzkin tree of size n, $n \to \infty$, which have an expected distance of

$$3\,2^{p-1}$$

to their critical successors. □

Note that we assumed here that the critical successors of a critical node with Horton-Strahler number 1 are the corresponding leaves of the Motzkin tree. From the fact that the total number of critical nodes changed from $\frac{4}{9}n$ to $\frac{1}{3}n$ by disregarding the leaves of a Motzkin tree we can conclude that a Motzkin tree with n nodes has $\frac{1}{3}n$ leaves on the average.
Beside the adjustments which lead to Corollary 2, we also have another possibility to determine the behavior of Motzkin trees without changing our computation a lot.

If we turn to a setting where we use the number of leaves ▧ of the Motzkin tree as a second parameter within our result, then we are able to solve the problem of the overestimated number of critical nodes very easily. For that purpose, we consider $K(z, \frac{1}{2}z, vz)$, now assuming that it is a bivariate function i.e. v is no longer assumed to be a fixed parameter. Then it is possible to adjust the number of critical nodes by subtracting ℓ from the coefficient at $z^n v^\ell$. The determination of $[z^n v^\ell]$ can be done be means of the following theorem due to Bender and Richmond [2].

Theorem 6 *Let $\phi_n(x_1,\ldots,x_d) = \sum_{\substack{k_i \geq 0 \\ i \in \{1,\ldots,d\}}} a_n(k_1,\ldots,k_d)x_1^{k_1}\ldots x_d^{k_d}$ and let R be a compact subset of $(0,\infty)^d$. Suppose $\phi_n(x_1,\ldots,x_d)$ converges in an ε neighborhood $N(R,\varepsilon)$ of R and*

$$\phi_n(x_1,\ldots,x_d) \sim f(n)g(x_1,\ldots,x_d)\lambda(x_1,\ldots,x_d)^n$$

uniformly in $N(R,\varepsilon)$, where $g(x_1,\ldots,x_d)$ is uniformly continuous and $\lambda(x_1,\ldots,x_d)$ has uniformly continuous third-order partials. Suppose that the matrix

$$B(x_1,\ldots,x_d) = \left(\frac{\partial^2}{\partial s_i \partial s_j}\ln\lambda(e^{s_1},\ldots,e^{s_d})|_{e^{s_i}=x_i,\,1\leq i\leq d}\right)_{i,j=1,\ldots,d}$$

is nonsingular for $(x_1,\ldots,x_d) \in R$ and $\phi_n(x_1,\ldots,x_d)/\phi_n(|x_1|,\ldots,|x_d|) = o(n^{-d/2})$ if $(|x_1|,\ldots,|x_d|) \in N(R,\varepsilon)$ and $(x_1,\ldots,x_d) \notin N(R,\varepsilon)$. Then, if $(k_1,\ldots,k_d)/n = m(t_1,\ldots,t_d)$ for $m(x_1,\ldots,x_d) = \nabla\ln\lambda(e^{s_1},\ldots,e^{s_d})|_{e^{s_i}=x_i,\,1\leq i\leq d}$, has a solution $(t_1,\ldots,t_d) \in R$ we have

$$a_n(k_1,\ldots,k_d) \sim \phi_n(t_1,\ldots,t_d)t_1^{-k_1}\cdots t_d^{-k_d}/\sqrt{(2\pi n)^d \det(B(\ln t_1,\ldots,\ln t_d))}$$

uniformly for all such $(k_1,\ldots,k_d)$. □

We consider the ε neighborhood $|v-1|<\varepsilon$. In that case, by setting $x := 1$ and $u := \frac{1}{2}$, the asymptotic formulæ for $[z^n]$ given in the Theorems 1 to 5 provide the asymptotic expansion $f(n)g(v)\lambda(v)^n$ for the different generating functions considered. This can be concluded in the following way. Obviously, for $v = 1$ the dominant singularity $z_d(v)$ of the bivariate generating functions is equal to the dominant singularity $z_d = \frac{1}{3}$ of the univariate generating functions which result from setting v to 1. Thus, for $v = 1$, the theorems apply and provide the asymptotics. For v sufficiently close to 1, the dominant singularity $z_d(v)$ lies in a neighborhood of z_d and depends analytically on v. Therefore, by general properties of the $\mathcal{O}$-transfer method, the asymptotics given in the theorems provide the desired expansions. The reader is refered to [12] and [11] for details on this sort of reasoning. For all generating functions involved, $\lambda(v) = \frac{1}{z_d(v)} = 1 + 2\sqrt{v}$ holds. Furthermore, $m(t) = \sqrt{t}/(1+2\sqrt{t})$ with the solution $t = \ell^2/(2\ell - n)^2$, $2\ell < n$, for the equation $\ell/n = m(t)$. Thus the expected number of critical nodes in Motzkin trees with n nodes and ℓ leaves is asymptotically given by

$$\frac{[z^n]K(z, \frac{1}{2}z, vz)|_{v=t} - \ell}{[z^n]T(z, \frac{1}{2}z, vz)|_{v=t}}.$$

This procedure applied to the different parameters yields the following corollary:

Corollary 3 *On the assumption that all Motzkin trees of the same size (n, ℓ) (total number of nodes n, ℓ leaves) are equally likely, the average Horton-Strahler number of a Motzkin tree of size (n, ℓ) is asymptotically given by*

$$\log_4\left(2\pi^2 \rho n\right) - \frac{2+\gamma}{2\ln(2)} + \Delta(n), \rho := \frac{\ell}{n} < \frac{1}{2} \text{ fix}, n \to \infty.$$

For a fixed ratio $\rho := \frac{\ell}{n} < \frac{1}{2}$ the corresponding r-th moments are given by

$$2^{-r}\log_2^r(n),\ n \to \infty.$$

The average number of critical nodes is asymptotically given by

$$\frac{1}{3}\rho n + \frac{1}{12}\log_2(n),\ \rho := \frac{\ell}{n} < \frac{1}{2} \text{ fix},\ n \to \infty.$$

On the average, there are

$$4^{-p}\ell$$

critical nodes with fixed Horton-Strahler number p in a Motzkin tree of size (n, ℓ), $\frac{\ell}{n} < \frac{1}{2}$, $n \to \infty$ which have an expected distance of

$$2^{p-1}\frac{n}{\ell}$$

to their critical successors. □

The results for the average Horton-Strahler number and the average number of critical nodes are only valid for ρ fix. This is a consequence of regarding ω as a single variable while determining the Mellin-transform which implies a coupling of the variables x, u, v and z. The observation that a Motzkin tree with n nodes has $\frac{1}{3}n$ leaves on the average provides a link between Corollary 2 and Corollary 3 since the application of the equality $\ell = \frac{n}{3}$ makes it possible to transform the result of one Corollary into the corresponding result of the other.

3.3 General Unary-Binary Trees

We can generalize the ideas used to investigate the Motzkin trees in order to consider unary/binary trees with c_1 different types of unary nodes and c_2 different types of binary nodes. The average Horton-Strahler number of those trees has been determined in [10]. As an example, those trees can be considered as arithmetic expressions where we have $c_1 > 0$ different unary and $c_2 > 0$ different binary operators from which the expression can be constructed. By setting $x := c_2$, $u := \frac{1}{2}c_1$ and $v := 1$ (resp. $v := v$) we find:

Corollary 4 *Let c_1 (resp. c_2) be the number of different types of unary (resp. binary) nodes. On the assumption that all unary/binary trees of the same size n (total number of nodes) (resp. (n, ℓ) (total number of nodes n, ℓ leaves)) are equally likely, the average Horton-Strahler number of a tree of size n (resp. size (n, ℓ)) is asymptotically given by*

$$\log_4\left(\frac{2\pi^2\sqrt{c_2}n}{2\sqrt{c_2}+c_1}\right) - \frac{2+\gamma}{2\ln(2)} + \Delta(n),\, n \to \infty,$$

$$\left(\textit{resp. } \log_4\left(2\pi^2\rho n\right) - \frac{2+\gamma}{2\ln(2)} + \Delta(n), \rho := \frac{\ell}{n} < \frac{1}{2} \textit{ fix}, n \to \infty.\right)$$

In both cases and for a fixed ratio $\rho := \frac{\ell}{n} < \frac{1}{2}$ the corresponding r-th moments are given by

$$2^{-r}\log_2^r(n),\, n \to \infty.$$

The average number of critical nodes in a tree of size n (resp. in a tree of size (n, ℓ)) is asymptotically given by

$$\frac{1}{3}\frac{\sqrt{c_2}n}{2\sqrt{c_2}+c_1} + \frac{1}{12}\log_2(n),\, n \to \infty,$$

$$\left(\textit{resp. } \frac{1}{3}\rho n + \frac{1}{12}\log_2(n),\, \rho := \frac{\ell}{n} < \frac{1}{2} \textit{ fix}, n \to \infty.\right)$$

On the average, there are

$$\frac{\sqrt{c_2}n}{4^p(2\sqrt{c_2}+c_1)} \;\left(\textit{resp. } 4^{-p}\ell\right)$$

critical nodes with fixed Horton-Strahler number p in a tree of size n (resp. size (n, ℓ)), $n \to \infty$, which have an expected distance of

$$\frac{2^{p-1}(2\sqrt{c_2}+c_1)}{\sqrt{c_2}} \;\left(\textit{resp. } 2^{p-1}\frac{n}{\ell}\right)$$

to their critical successors. □

Note that the influence of c_2 and c_1 has cancelled out within our bivariate results related to critical nodes. Note further that we rediscover the results for Motzkin trees by setting $c_1 := c_2 := 1$. Our results for the expected Horton-Strahler number implies the conjecture that the average number of leaves in a general unary/binary tree is asymptotically given by $\frac{\sqrt{c_2}n}{2\sqrt{c_2}+c_1}$. This conjecture proves to be true by obvious computations based on the generating function $T(x, u, v)$.

3.4 Combinatorial Tries

The family of binary trees called $\mathcal{C}$-tries has been introduced in [22] as a combinatorial model for the trie data structure. The family of $\mathcal{C}$-tries is symbolically defined in (2). If we consider a leaf $\square$ to store a key and a leaf $\blacksquare$ to represent a NIL pointer then all possible tree structures of a digital trie are resembled. The motivation for a combinatorial model of tries lies e.g. in the application of tries to the compression of blockcodes (see [24] for details).
In order to derive results for $\mathcal{C}$-tries we have to introduce a new variable y to mark non-empty leaves (which are those storing the keys of the trie). According to (2) leaves of an internal node of type 0 must both store a key, while the leaf of an internal node of type 1 may either store a key or may be a NIL pointer. Thus we have to set $x := 1$, $u := 1 + y$ and $v := y^2$. We use Theorem 6 to compute an asymptotic for the coefficients at $[z^n y^\ell]$. Again, considering the ε-neighborhood $|y - 1| < \varepsilon$ implies that the asymptotics given in the Theorems 1 to 5 provide the appropriate expansions. We find that $\lambda(y)$ is given by $2 + 4y$ and thus $m(t) = 2t/(2t + 1)$ holds. The solution of $m(t) = \ell/n$ is given by $t = \frac{\ell}{2(n-\ell)}$, $\frac{\ell}{n} < 1$. The application of Theorem 6 finally yields:

Corollary 5 *On the assumption that all $\mathcal{C}$-tries of size (n, ℓ) (n internal nodes, ℓ keys) are equally likely the average Horton-Strahler number of a tree of size (n, ℓ) is asymptotically given by [23]*

$$\log_4\left(2\pi^2 \rho n\right) - \frac{2 + \gamma}{2\ln(2)} + \Delta(n),\ \rho := \frac{\ell}{n} < 1 \text{ fix},\ n \to \infty.$$

For a fixed ratio $\rho := \frac{\ell}{n} < 1$ the corresponding r-th moments are given by

$$2^{-r} \log_2^r(n),\ n \to \infty.$$

The average number of critical nodes in a $\mathcal{C}$-trie of size (n, ℓ) is asymptotically given by

$$\frac{1}{3}\rho n + \frac{1}{12}\log_2(n), \rho := \frac{\ell}{n} < 1 \text{ fix},\ n \to \infty.$$

On the average there are

$$4^{-p}\ell$$

critical nodes with Horton-Strahler number p in a tree of size (n, ℓ), $n \to \infty$, which have an expected distance of

$$2^p \frac{n}{\ell}$$

to their critical successors. $\square$

Note that univariate results (i.e. results where only the number of internal nodes determines the size) for $\mathcal{C}$-tries cannot be concluded directly from our results for general unary/binary trees by specific choices for c_1 and c_2, since there is no shift in the Horton-Strahler number and no overestimated number of critical nodes in the case of $\mathcal{C}$-tries.

4 Discussion of the Results

Obviously we could interpret all our results with respect to the different applications of the Horton-Strahler number mentioned in the introduction. However, this is not the aim of this section. Here we want to discuss conclusions that can be drawn with respect to general properties of the parameters and to the different families of trees

considered.
If we take a look at the results of the previous section we find the following similarities:

- The average Horton-Strahler number is always given by $\log_4(2\pi^2\mathcal{L}) - \frac{2+\gamma}{2\ln(2)} + \Delta(n)$ for $\mathcal{L}$ the expected asymptotical number of leaves in a tree of size n or in the case of bivariate results $\mathcal{L}$ the number of leaves specified by some parameter.
- The r-th moment is given by $2^{-r}\log_2^r(n)$ in all cases.
- The average total number of critical nodes is given by $\frac{1}{3}\mathcal{L} + \frac{1}{12}\log_2(n)$ for $\mathcal{L}$ as given in the first item.
- The average number of critical nodes with Horton-Strahler number p is given by $4^{-p}\mathcal{L}$ for $\mathcal{L}$ as given in the first item.
- The expected distance between a critical node with Horton-Strahler number p and its critical successors with mark $p-1$ is either $2^{p-1}\frac{n}{\mathcal{L}}$ for families of trees with a shifted Horton-Strahler number or $2^p\frac{n}{\mathcal{L}}$ otherwise.

Some of these observations can be explained by means of our theorems. For example, the quotient $\frac{[z^n]M^{(r)}(xz,uz,vz)}{[z^n]T(xz,uz,vz)}$ simplifies to $2^{-r}\log_2^r(n)$ for arbitrary choices of x, u and v. Thus every family of trees which can be analyzed by our approach posseses this asymptoical representation of the r-th moments. Furthermore, $\frac{[z^n]K_{p+1}(xz,uz,vz)}{[z^n]K_p(xz,uz,vz)} = \frac{1}{4}$ holds which explains the factor 4^{-p} within the expected number of critical nodes with Horton-Strahler number p. Additionaly this constant ratio implies that the number of critical nodes of any Horton-Strahler number and thus the total number of critical nodes only depends on the number k_1 of critical nodes with Horton-Strahler number 1. For all families of trees considered here, k_1 seems to be given by $\frac{1}{4}$ times the (expected) number of leaves of the tree. The fact that the number of critical nodes is quatered each time we increase the Horton-Strahler number by one also provides additional information on the internal structure of the trees. Since each critical node with Horton-Strahler number p cannot have more than 2 critical sucsessors with Horton-Strahler number $p-1$ we can conclude that one half of the critical nodes with Horton-Strahler number $p-1$ are successors of critical nodes with a Horton-Strahler number greater than p. This sort of self similarity seems to start at the very beginning, i.e. for $p=1$, since on the average only one half of the leaves *create* a critical node with Horton-Strahler number 1. There is a third invariant based on our general results. For each possible choice of x, u and v we find that $\frac{[z^n]KD_p(xz,uz,vz)}{[z^n]T(xz,uz,vz)}$ is given by $2^{-p}n$. Therefore, on the average the accumulated length of the critical paths for all critical nodes with Horton-Strahler number p within a tree of size n is the same for all families of trees that can be considered within our model.
It remains to mention that obviously not all substitutions are allowed. For example, it is not possible to set one of the variables in $\{x,u,v\}$ to zero since we assumed that each internal node contributes to the size of the tree when computing the asymptotics of the Theorems 1 to 5. Thus a value of zero for one of the variables would lead to inconsistencies with respect to the size. However, it is possible to derive corresponding results by considering $[z^n]$ for appropriate arguments of our generating functions (e.g. $R_p(xz,u,vz)$ instead of $R_p(xz,uz,vz)$). Furthermore, since there is no recomputation of the Horton-Strahler number it is not possible to apply substitutions which correspond to the insertion of binary tree structures which would imply a change of the Horton-Strahler number of the original tree. For example, to set x to $\frac{1-\sqrt{1-4z}}{2z}$ corresponds to the substitution of an internal node

of type 2 by any arbitrary binary tree and thus to a situation where the Horton-Strahler number of the resulting tree might not fit with the Horton-Strahler number of the initial tree considered by our generating functions. However, by properties of the Horton-Strahler numbers, a substitution of a single node by a linear list is possible.

5 Conclusions

In this paper we have introduced a method of how to analyze parameters related to the Horton-Strahler number of binary trees in a unified way such that one detailed computation for each parameter is sufficient to get results for different families of trees. This is a slight improvement since so far it was standard to perform a dedicated analysis of comparable complexity for each family of trees. The generating functions presented here proved to be useful to solve an old problem related to the secondary structure of single stranded nucleic acids. In [25] it is shown how $R_p(x,u,v)$ can be used to asymptotically enumerate the number of secondary structures of order p built from n bases. Again, appropriate substitutions for x, u and v turned out to be the key to the solution. However, the future will show whether the method presented can also be applied to other types of problems. Furthermore, it was possible to conclude some invariants that are fulfilled by each family of trees that can by analyzed by means of our approach. We want to conclude this paper by remarking that the method presented is general enough to handle nodes of higher degree. It was due to the generalized Horton-Strahler numbers that e.g. nodes with three successors were not considered in our computations since in such a case there is no closed form solution known for the resulting recursive representation of the generating functions involved.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, 1970

[2] E. A. Bender and L. B. Richmond, *Central and Local Limit Theorems Applied to Asymptotic Enumeration II: Multivariate Generating Functions*, Journal of Combinatorial Theory, Series A **34**, 255-265, 1983.

[3] N. G. de Bruijn, D. E. Knuth and S.O. Rice, *The Average Height of Planted Plane Trees*, Graph Theory and Computing (R. C. Read, ed.), Academic Press, 1972

[4] L. Comtet, *Advanced Combinatorics*, D. Reidel, 1974.

[5] L. Devroye and P. Kruszewski, *On the Horton-Strahler Number for Random Tries*, Theoretical Informatics and Applications **30**, 443-456, 1996

[6] A. P. Ershov, *On Programming of Arithmetic Operations*, Communications of the ACM **1**, 3-6, 1958

[7] P. Flajolet, J.C. Raoult and J. Vuillemin, *The Number of Registers required for Evaluating Arithmetic Expressions*, Theoretical Computer Science **9**, 99-125, 1979

[8] P. Flajolet, *Analyse d'Algorithmes de Manipulation d'arbres et de fichier*, Cahiers du BURO **34-35**, 1-209, 1981

[9] P. Flajolet and A. Odlyzko, *Singularity Analysis of Generating Functions*, SIAM J. Disc. Math. **3**, No. 2, 216-240, 1990

[10] P. Flajolet and H. Prodinger, *Register Allocation for Unary-Binary Trees*, SIAM J. Comput. **15**, 629-640, 1986

[11] P. Flajolet and R. Sedgewick, *The Average Case Analysis of Algorithms: Multivariate Asymptotics and Limit Distributions*, INRIA rapport de recherche **3162** (1997).

[12] P. Flajolet and M. Soria, *General Combinatorial Schemas: Gaussian Limit Distributions and Exponential Tails*, Discrete Mathematics **114**, 159-180, 1993

[13] P. Flajolet, X. Gourdon and P. Dumas, *Mellin transforms and asymptotics: Harmonic sums*, Theoretical Computer Science **144**, 3-58, 1995

[14] J. Françon, *Sur le nombre de registres nécessaires à l'évaluation d'une expression arithmétique*, R.A.I.R.O. Theoretical Informatics and Applications **18**, 355-364, 1984

[15] I. Herman, M. Delest and G. Melançon, *Tree visualisation and navigation clues for information visualisation*, Computer Graphics Forum **17**, 153-165, 1998

[16] R. E. Horton, *Erosioned development of systems and their drainage basins, hydrophysical approach to quantitative morphology*, Bull. Geol. Soc. of America **56**, 275-370, 1945

[17] R. Kemp, *The Average Number of Registers Needed to Evaluate a Binary Tree Optimally*, Acta Informatica **11**, 363-372, 1979

[18] R. Kemp, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley-Teubner Series in Computer Science, 1984

[19] D. E. Knuth, *The Art of Computer Programming, Vol. 3, Sorting and Searching*, Addison Wesley, 1998.

[20] A. Meir and J. W. Moon, *On the Altitude of Nodes in Random Trees*, Can. J. Math. **30**, 997-1015, 1978

[21] A. Meir, J. W. Moon and J. R. Pounder, *On the Order of Random Channel Networks*, SIAM J. Alg. Disc. Math. **1**, 25-33, 1980

[22] M. E. Nebel, *The Stack-Size of Tries, A Combinatorial Study*, Theoretical Computer Science, to appear

[23] M. E. Nebel, *On the Horton-Strahler number for Combinatorial Tries*, R.A.I.R.O. Theoretical Informatics and Applications **34**, 279-296, 2000

[24] M. E. Nebel, *The Stack-Size of Combinatorial Tries Revisited*, submitted, 2001

[25] M. E. Nebel, *Combinatorial Properties of RNA Secondary Structures*, Frankfurter Informatik-Berichte 2/01, Institut für Informatik, Johann Wolfgang Goethe-Universität, Frankfurt a. M., 2001

[26] H. Prodinger, *Die Bestimmung gewisser Parameter von binären Bäumen mit Hilfe analytischer Methoden*, Zahlentheoretische Analysis, Lecture Notes in Mathematics **1114**, 118-133, 1985.

[27] H. Prodinger, *Some Recent Results on the Register Function of a Binary Tree*, Annals of Discrete Mathematics **33**, 241-260, 1987

[28] A. N. Strahler, *Hypsometric (area-altitude) analysis of erosonal topology*, Bull. Geol. Soc. of America **63**, 1117-1142, 1952

[29] M. S. Waterman, *Secondary Structure of Single-Stranded Nucleic Acids*, Advances in Mathematics Supplementary Studies **1**, 167-212, 1978

[30] M. Vauchaussade de Chaumont, *Nombre de Strahler des arbres, langages algébriques et dénombrement de structures secondaires en biologie moléculaire*, Thèse 3'eme cycle, Université de Bordeaux I, Mai 1985

[31] X. G. Viennot, G. Eyrolles, N. Janey and D. Arques, *Combinatorial analysis of ramified patterns and computer imagery of trees*, Proc. SIGGRAPH'89, Computer Graphics **23**, 31-40, 1989

[32] X. G. Viennot, *Trees Everywhere*, Proc. CAAP'90, LNCS **431**, 18-41, 1990