The Stack-Size of Combinatorial Tries Revisited

Markus E. Nebel

Discrete Mathematics and Theoretical Computer Science **5**, 2002, 1–16

The Stack-Size of Combinatorial Tries Revisited

Markus E. Nebel

Johann Wolfgang Goethe-Universität, Fachbereich Biologie & Informatik, Frankfurt am Main, Germany

received Jun 15, 2001, accepted Jan 24, 2002.

In the present paper we consider a generalized class of extended binary trees in which leaves are distinguished in order to represent the location of a key within a trie of the same structure. We prove an exact asymptotic equivalent to the average stack-size of trees with α internal nodes and β leaves corresponding to keys; we assume that all trees with the same parameters α and β have the same probability. The assumption of that uniform model is motivated for example by the usage of tries for the compression of blockcodes. Furthermore, we will prove asymptotics for the r-th moments of the stack-size and we will show that a normalized stack-size possesses a theta distribution in the limit.

Keywords: Tries, Blockcodes, Stack-Size, Analytic Combinatorics.

1 Introduction

A binary tree is a rooted ordered tree where each node has at most two descendants. A digital *trie* is a binary tree which is used to store the set of keys $K = \{k_1, \ldots, k_n\}$ in the following manner: The prefixes of the binary representations of the keys k_i, $1 \leq i \leq n$, considered as strings of 0's and 1's, are used to navigate through the tree; for each 0 (resp. 1) a left (resp. right) branch is used in order to go to the next level of the tree. The trie T for the set of keys K is the smallest tree for which all these paths are different. Thus for $\mathcal{P}(K)$ the set of prefixes of the keys in K, the set $\mathsf{INIT}(K) := \{u \in \mathcal{P}(K) : |\{v \in K : v = u \cdot \{0,1\}^\star\}| = 1\}$ contains exactly the words that correspond to a path from the root of T to one of the leaves. Note that T might have internal nodes with only one (left or right) successor. Those internal nodes are avoided by the Patricia algorithm (see [20] for details on the implementation) in order to achieve more compact trees. In the same manner it is possible to construct an m-ary trie (or m-ary Patricia tries) from data which possess an m-ary representation, like character-strings. For details on the implementation see [20] and [13].

Assuming that the set of keys K is a set of random integers and we use their binary representations to navigate through the trie, we observe that it is much more likely to get a trie which is a balanced tree than a trie which is a linear list. The reason for this fact is that a linear structure of length n is only implied by at least two keys with a common prefix in their binary representation of length n; for two random integers the probability for such a prefix decreases like 2^{-n}. Thus, a trie profits from properties of the input-data. For the mathematical analysis of tries, the probability model known as *Bernoulli model* (see e.g. [22]), could be used to take those phenomena into account. Parameters which were consider in this model are for example the height of tries, the external path length of tries, the depth of leaves (keys) in tries and the

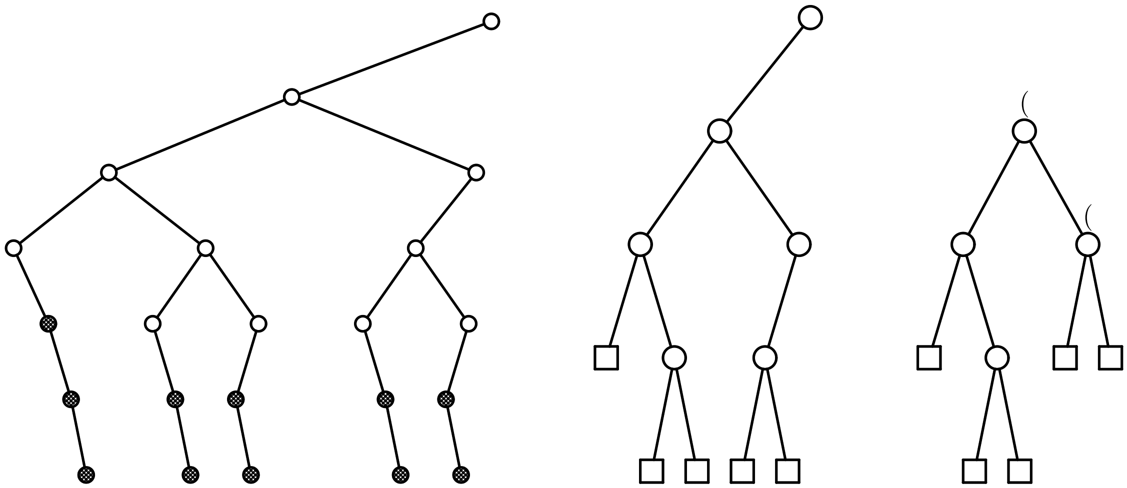

Fig. 1: The code tree of the semi Dyck-language of length 6 (left tree), its compact representation as a trie (tree in the middle) and the resulting Patricia trie (right tree).

size (number of internal nodes) of a trie (see e.g. [4, 5, 7, 14, 18, 26, 29, 30]). Note, that there is also a rich literature considering Patricia tries in the Bernoulli model (see e.g. [19, 25, 31] and the references given there).

Tries also appear in other fields of computer science. For example, it is common practice to represent a binary blockcode of length n by a 0-balanced binary tree of height n, called *code tree*. Each leaf of this tree corresponds to exactly one code word which is generated by the concatenation of the edge labels (0 for an edge to a left son, 1 for an edge to a right son) on the path from the root to the leaf. We can get a compressed representation (a compressed code) by successively deleting such leaves in the tree that do not have a brother. The resulting tree possesses the structure of a trie. The external path length of such a trie is related to the cost for ranking the encoded objects [21]. The preorder traversal of that trie yields a lexicographical enumeration of all code words by writing out the related string each time a leaf is reached. An additional compression can be obtained by the application of the Patricia algorithm which deletes the linear lists inside a trie. In that case the deleted edge labels are attached to the corresponding internal nodes. Figure 1 shows the code tree for all semi Dyck words of length 6, the corresponding compressed representation as a trie and the resulting Patricia trie. In order to prepare this example, an edge to a left (resp. right) son was used to represent an opening (resp. closing) bracket. With respect to the Bernoulli model, the trie of Figure 1 (tree in the middle) is rather unlikely. However, there is no reason why this trie should have a smaller probability than any other trie which results from the compression of a code tree. Thus, in the context of binary blockcodes, the assumption of a uniform distribution seems to be reasonable.

In the present paper we therefore have a look at the combinatorics of digital tries that so far has not been studied a lot. We will use the fact that an internal node v of a trie might have an empty successor (a NIL-pointer) if and only if the other successor of v is an internal node (otherwise v would be a leaf or the successor of v could be deleted to achieve a better compression of the code (v could be used to store the data of its succeeding leaf)). Following the approach presented in [11], a digital trie possesses the

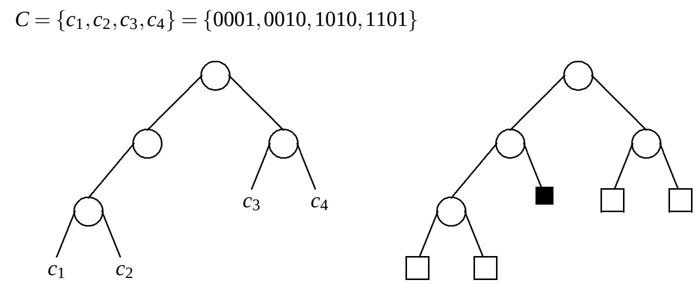

Fig. 2: An example for a set of code words (keys) C, the resulting trie and the corresponding generalized extended binary tree.

structure of an extended binary tree with colored leaves. The appropriate class of extended binary trees has been introduced in [23] where a *generalized extended binary tree* was defined as a binary tree with colored leaves; leaves are either colored black (represented as ■) or white (represented as □) in such a way that each black leaf is the brother of an internal node. If we now assume a black leaf to represent an empty position (a NIL-pointer) and a white leaf to represent a code word (a key), exactly those tree-structures that can be generated by the compression of code trees (by the trie algorithm) are resembled. An example for that correspondence can be found in Figure 2, where C can be considered either as a set of code words of a blockcode of length 4 or as a set of binary integers for which the trie implements a dictionary.

For the sake of simplicity we will use the term $\mathcal{C}$-tries to denote the class of generalized extended binary trees. A $\mathcal{C}$-trie with α internal nodes and β white leaves will be called (α,β)-trie; we will call (α,β) the size of such a tree. Note that $2 \leq \beta \leq \alpha+1$ must hold. Furthermore, the ratio $\rho := \frac{\alpha}{\beta}$, which will show up within our results, is a measure for the *distance* of the trie to the corresponding Patricia trie and therefore related to the utilization of the tree. In connection with blockcodes it thus can be interpreted as the degree of redundancy of the code. A Patricia trie without redundancy always fulfills $\beta = \alpha+1$ which implies a lower bound of ρ for all (α,β)-tries.
The main interest of this article is the stack-size of uniform random $\mathcal{C}$-tries (i.e. we assume that all (α,β)-tries have the same probability) where the stack-size $s(T)$ of a tree T is defined as follows:

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

Here $T.l$ (resp. $T.r$) denotes the left (resp. right) subtree of T. When traversing T by means of an optimized[†] recursive procedure in preorder, $s(T)$ denotes the recursion-depth of the traversal; without the application of the optimization the height of T corresponds to the recursion-depth. We will show results on the average stack-size of uniform random $\mathcal{C}$-tries, on the related higher moments and on the limiting distribution. The results presented will depend on both, the number of internal nodes α and the number of white leaves (keys) β. So far, the stack-size of $\mathcal{C}$-tries has only been considered with respect to the number of internal nodes (see [23]) disregarding the number of white leaves (keys). Up to now, there were no results on higher moments and on the distribution.

[†] Applying a technique known as ‘end recursion removal’ the recursion-depth can be reduced. See [28] for details.

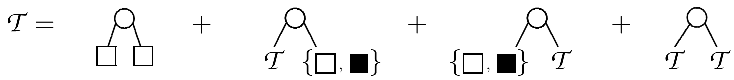

Fig. 3: The construction of a $\mathcal{C}$-trie with at least one internal node. The set $\{\square, \blacksquare\}$ represents the possibility of choosing either a white or a black leaf.

Remark: It is also possible to interpret the stack-size of a tree in a completely different manner. If we think of T as a syntax-tree which represents an arithmetic expression $\mathcal{E}$, then $s(T)$ specifies the number of cells on a stack that are needed to evaluate $\mathcal{E}$ by means of a simple traversal algorithm (see [17, pp. 132] for details). With respect to this notion it is possible to think of a $\mathcal{C}$-trie as a special kind of arithmetic expression built of unary and binary operators. In this context the assumption of a uniform probability distribution is quite natural, too. From a strict combinatorial point of view the stack-size can also be considered as the *right-height* of the tree as it equals the notion of height, when only right edges, and not all edges, contribute to the height.

In the sequel the notation $[x^\alpha y^\beta] f(x,y)$ is used to represent the coefficient at $x^\alpha y^\beta$ in the expansion of $f(x,y)$ at $(x,y) = (0,0)$.

2 The Average Stack-Size

In this section we will derive explicit formulæ for the average stack-size of $\mathcal{C}$-tries. For that purpose we need the total number $T_{\alpha,\beta}$ of (α,β)-tries.

Lemma 1 *The number $T_{\alpha,\beta}$ of $\mathcal{C}$-tries of size (α,β) is given by*

$$T_{\alpha,\beta} = \frac{2^{\alpha-\beta+1}}{\beta}\binom{2\beta-2}{\beta-1}\binom{\alpha-1}{\alpha-\beta+1}.$$

Proof: Let x mark an internal node and y mark a white leaf. The construction process for a $\mathcal{C}$-trie as shown in Figure 3 translates directly into the equation

$$T(x,y) = xy^2 + (2xy + 2x)T(x,y) + xT^2(x,y),$$

for $T(x,y)$ the ordinary generating function of $\mathcal{C}$-tries. Therefore, we find that $T(x,y) = \frac{1-2xy-2x-\sqrt{1-4\kappa}}{2x}$ with $\kappa = x - x^2 + xy - 2x^2y$ holds. Now we add y in order to take the tree with zero internal nodes into account. This yields

$$T(x,y) = \frac{1-2x-\sqrt{1-4\kappa}}{2x} = \frac{1/2-x}{x} - \frac{1-2x}{2x}\sqrt{1-4xy(1-2x)^{-1}}. \tag{1}$$

Note that the term $\frac{1/2-x}{x}$ does not contribute to $T_{\alpha,\beta}$ since it cancels when we expand the squareroot. By expanding the other term of $T(x,y)$ in the usual way we find the statement of the lemma. $\square$

Next, we investigate the class of those $\mathcal{C}$-tries that possess the same stack-size. To determine the ordinary generating function $A_k(x,y)$ of $\mathcal{C}$-tries that could be traversed with stack-size less than $k+1$, we have to distinguish between the cases given in Figure 4. Denoting the ordinary generating function of $\mathcal{C}$-tries with

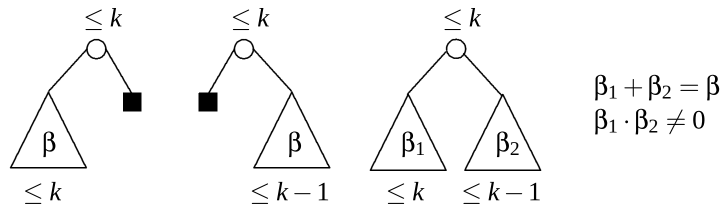

Fig. 4: All possible decompositions of a $\mathcal{C}$-trie with β white leaves and a stack-size of at most k. The number inside a triangle corresponds to the number of white leaves it has to possess, the number below a triangle determines the stack-size of the subtree represented by it.

a stack-size of at most k and j white leaves by $L_{k,j}(x,y)$, these cases translate into the following set of equations:

$$\begin{aligned}
L_{k,1}(x,y) &= y, k \geq 1,\\
L_{1,j}(x,y) &= 0, j \geq 2,\\
L_{k,j}(x,y) &= xL_{k,j}(x,y) + xL_{k-1,j}(x,y) + x \sum_{\substack{\beta_1+\beta_2=j\\ \beta_1\cdot\beta_2\neq 0}} L_{k,\beta_1}(x,y)L_{k-1,\beta_2}(x,y)\\
&= \left(xL_{k-1,j}(x,y) + x \sum_{\substack{\beta_1+\beta_2=j\\ \beta_1\cdot\beta_2\neq 0}} L_{k,\beta_1}(x,y)L_{k-1,\beta_2}(x,y) \right)(1-x)^{-1}.
\end{aligned}$$

Now $A_k(x,y) = \sum_{j\geq 1} L_{k,j}(x,y)$ holds and thus

$$\begin{aligned}
A_1(x,y) &= y,\\
A_k(x,y) &= -1 + \frac{2xy+x-y-1}{-1+x+xA_{k-1}(x,y)}, k \geq 2.
\end{aligned}$$

We use a result of [23] where a general solution for continued fractions of the pattern $C_k(x) = -1 + \frac{c_1}{c_2+xC_{k-1}(x)}$, $C_1(x) = c_3$, was given. Setting $c_1 := 2xy+x-y-1$, $c_2 := -1+x$ and $c_3 := y$ implies the following representation for $A_k(x,y)$:

$$A_k(x,y) = \frac{y - xyS_k(u)}{1-(x+xy)S_k(u)} \tag{2}$$

with $S_k(u) := \frac{1-u^{k-1}}{1-u^k}(1+u)$, $u := \frac{(1-\varepsilon)}{(1+\varepsilon)}$ and $\varepsilon := \sqrt{1-4\kappa}$, $\kappa := x - x^2 + xy - 2x^2y$. Our next task is to determine $[x^\alpha y^\beta]A_k(x,y)$ for which we use the following lemma:

Lemma 2 ([16]) *Let $\mathcal{S}_k(x) := \frac{1-\vartheta^k}{1-\vartheta^{k+1}}(1+\vartheta)$ with $\vartheta := (1-\sqrt{1-4x})/(1+\sqrt{1-4x})$ be the generating function of those extended binary trees that could be traversed with at most k cells of stack. Then for $i \geq 1$*

$$\begin{aligned}
\mathcal{S}^i_{k-1}(x) &= \sum_{n\geq 0} x^n \sum_{\lambda\geq 0}\sum_{h\geq 0} (-1)^\lambda \binom{i}{\lambda}\binom{i-1+h}{i-1}\\
&\quad \times \left[\binom{2n+i-1}{n-(k-1)\lambda-kh} - \binom{2n+i-1}{n-(k-1)\lambda-kh-1} \right].
\end{aligned}$$

□

In order to determine the coefficient in question we expand the right-hand side of (2) into

$$y\sum_{i\geq 0}\sum_{j\geq 0}\binom{i}{j}x^iy^jS_k^i(u)-xy\sum_{i\geq 0}\sum_{j\geq 0}\binom{i}{j}x^iy^jS_k^{i+1}(u)$$

and use Lemma 2 to represent the powers of function S_k. Then it is easy to extract the coefficient $[x^\alpha y^\beta]$ in the resulting formula. We obtain:

$$\begin{aligned}S_{k,\alpha,\beta}&:=[x^\alpha y^\beta]A_k(x,y)=\\&\sum_{i\geq 0}\sum_{j\geq 0}\binom{i}{j}\sum_{m\geq 0}\binom{m}{\beta-1-j}\sum_{v\geq 0}\binom{m-\beta+1+j}{v}(-1)^v(-2)^{\alpha-i-m-v}\\&\times\left[\binom{\beta-1-j}{\alpha-i-m-v}\varphi(i,m,k)-\binom{\beta-1-j}{\alpha-i-m-v-1}(-2)^{-1}\varphi(i+1,m,k)\right]\end{aligned}$$

for $\alpha\geq 1$ and $S_{k,\alpha,\beta}=\delta_{\beta,1}$ for $\alpha=0$. Here $\varphi(i,m,k)$ is defined as

$$\begin{aligned}&\varphi(i,m,k):=\\&\sum_{\lambda\geq 0}\sum_{\ell\geq 0}(-1)^\lambda\binom{i}{\lambda}\binom{i-1+\ell}{i-1}\left(\binom{2m+i-1}{m-(k-1)\lambda-k\ell}-\binom{2m+i-1}{m-(k-1)\lambda-k\ell-1}\right).\end{aligned}$$

Applying some further simplifications leads to the following lemma:

Lemma 3 *The number $S_{k,\alpha,\beta}$ of (α,β)-tries with a stack-size $\leq k$ is $\delta_{\beta,1}$ for $\alpha=0$ and*

$$\begin{aligned}S_{k,\alpha,\beta}\;=\;&\sum_{i\geq 1}\sum_{j\geq 0}\frac{j}{i}\binom{i}{j}\sum_{m\geq 0}\binom{m}{\beta-1-j}\\&\times\sum_{v\geq 0}\binom{m-\beta+1+j}{v}(-1)^v(-2)^{\alpha-i-m-v}\binom{\beta-1-j}{\alpha-i-m-v}\varphi(i,m,k)\end{aligned}$$

for $\alpha\geq 1$. □

Now, for $S_{0,\alpha,\beta}=0$, the average stack-size is given by the quantity

$$T_{\alpha,\beta}^{-1}\sum_{1\leq k\leq\alpha+1}k(S_{k,\alpha,\beta}-S_{k-1,\alpha,\beta})=(\alpha+1)-|T_{\alpha,\beta}|^{-1}\sum_{1\leq k\leq\alpha}S_{k,\alpha,\beta}$$

for $T_{\alpha,\beta}$ as given in Lemma 1. Using our representation of $S_{k,\alpha,\beta}$ and performing a lengthy computation similar to that in [23] we find:

Theorem 1 *Under the assumption of the uniform model, the average stack-size of a (α,β)-trie is equal to*

$$\begin{aligned}&1+\left\{\sum_{i\geq 1}\sum_{j\geq 0}\frac{j}{i}\binom{i}{j}\sum_{m\geq 0}\binom{m}{\beta-j-1}\sum_{v\geq 0}\binom{m-\beta+j+1}{v}\binom{\beta-j-1}{\alpha-i-m-v}\right.\\&\times(-1)^v(-2)^{\alpha-i-m-v}\sum_{\lambda\geq 0}(-1)^{\lambda+1}\binom{i}{\lambda}\sum_{\ell\geq 1}\left[\binom{2m+i-1}{m-\ell+\lambda}-\binom{2m+i-1}{m-\ell+\lambda-1}\right]\\&\left.\times\sum_{d\mid\ell}\binom{i-1+d-\lambda}{i-1}\right\}T_{\alpha,\beta}^{-1}\end{aligned}$$

with $T_{\alpha,\beta}$ as given in Lemma 1. □

This result does not give us a lot of information about the behavior of the average stack-size since it is too complicated. Thus we have to derive an asymptotic equivalent.

3 Asymptotical r-th Moments and Limiting Distribution

In this section we will consider asymptotics for all moments of the stack-size and we will prove that a normalized stack-size possesses the theta distribution in the limit. Similar observations concerning the height of trees can be found in [9]. The usage of moments about the origin proved to be the method of choice in our context.

Please recall that $u = \frac{(1-\varepsilon)}{(1+\varepsilon)}$ for $\varepsilon = \sqrt{1-4\kappa}$ and $\kappa = x - x^2 + xy - 2x^2y$. We define $A_0(x,y)$ to be zero and set

$$a := \frac{(1-x-xu)y}{1-(x+xy)(1+u)},\ b := \frac{(u(x-1)+x)y}{u(1-(x+xy)(1+u))},\ c := \frac{u-(x+xy)(1+u)}{u(1-(x+xy)(1+u))}$$

and

$$\overline{A}_p(x,y) = \frac{acu^p}{1-cu^p} + \frac{bu^p}{1-cu^p},\ p \geq 1.$$

By using the relation $A_p(x,y) - \overline{A}_p(x,y) = a$ (since $a = y + T(x,y)$ holds, $-\overline{A}_p(x,y)$ is the generating function of $\mathcal{C}$-tries with a stack-size greater than p) one easily realizes that the generating function of the r-th moment

$$\sum_{p\geq 1} p^r \left(A_p(x,y) - A_{p-1}(x,y)\right)$$

for $r \geq 1$ can be rewitten as

$$a - \sum_{p\geq 1}\sum_{j\geq 1}\binom{r}{j} p^{r-j}\overline{A}_p(x,y). \tag{3}$$

We have to consider the function $M_\nu(x,y) := -\sum_{p\geq 1} p^\nu \overline{A}_p(x,y)$. Using (2) we find that

$$M_\nu(x,y) = -\left(a+\frac{b}{c}\right)\underbrace{\sum_{n\geq 1} u^n \sum_{d\mid n} c^d \left(\frac{n}{d}\right)^\nu}_{=:\varsigma(u)}$$

holds. We will use the O-transfer method as presented in [10] together with the saddle point method (see [12]) to obtain an asymptotic for the coefficient $[x^\alpha y^\beta]M_\nu(x,y)$. For this purpose we consider $M_\nu(x,y)$ as a function in x with complex parameter y and determine the dominant singularity $x_0(y)$ of $M_\nu(x,y)$. It is sufficient to restrict our attention to y with $|y| = 1/(2(\rho-1))$ for a fixed ratio $\rho := \alpha/\beta > 1$. Obviously, $M_\nu(x,y)$ converges (and thus is an analytic function) if $|u| < 1$ and $|c| \leq 1$ hold. Conversely, for $u = c = 1$ $M_\nu(x,y)$ becomes divergent and thus singular. The latter occurs for $x = x_0(y) := 1/(2+4y)$. In order to apply the O-transfer method we must be able to extend the function analytically beyond its disc of convergence. More precisely the function must be Δ-analytic, i.e. it must be analytic in the open domain

$$\Delta(\phi,R) := \{z \mid |z| < R, z \neq z_0(y), |\arg(z-z_0(y))| > \phi\}$$

for some $R > |z_0(y)|$ and $0 < \phi < \frac{\pi}{2}$. It is not hard to show that $|u| < 1$ for $\Re(x) < \frac{1+|y|}{2+4|y|}$ and $\Im(x) \neq 0$ and that $|u| < 1$ for $\Im(x) = 0$ and $\Re(x) < \frac{1}{2+4|y|}$. Furthermore, $|c| \leq 1$ as long as $\Re(x) < \frac{1+|y|}{2+4|y|}$. Thus it is obvious that we can choose R and ϕ such that $M_\nu(x,y)$ is Δ-analytic. Next we need an expansion of $M_\nu(x,y)$ at $x_0(y)$ which we will derive by means of the Mellin summation technique. We set $u = \exp(-t)$ within ς, then we compute the Mellin transform of the resulting sum with respect to variable t (the factor $-(a+b/c)$ will be considered later). We find

$$\varsigma(e^{-t}) \xrightarrow{\mathcal{M}} \Gamma(s) \sum_{n\geq 1} n^{-s} \sum_{d|n} c^d \left(\frac{n}{d}\right)^\nu.$$

Due to properties of the Dirichlet convolution and the application of the exp / log trick we find a simpler representation of the transform, namely

$$\Gamma(s)\zeta(s-\nu)\sum_{i\geq 0} \ln^i(c)\zeta(s-i)/i!.$$

Before we can use the Mellin summation formula given in [8] we have to set $u = \exp(-t)$ within $\ln(c)$ also. An expansion of the resulting expression at $t = 0$ yields the appropriate approximation

$$\ln(c) = -2\frac{x(1+y)}{1-2x-2xy}t + O(t^3).$$

Now, according to the methodology, an expansion of $\varsigma(e^{-t})$ at $t = 0$ is given by the sum of the residues of $t^{-s}\Gamma(s)\zeta(s-\nu)\sum_{i\geq 0}\left(-2\frac{x(1+y)}{1-2x-2xy}t\right)^i \zeta(s-i)/i!$. For each i fixed and $\nu \geq 1$ the most significant contribution of that sum of residues is of the order $O(t^{-(\nu-i+1)})$. Thus, for the leading term, only the summand for $i = 0$ contributes. In that way we find that the most significant term of the expansion of our sum is given by

$$\Gamma(\nu+1)\zeta(\nu+1)t^{-(\nu+1)}, \nu \geq 1.$$

Contributions of lower significance will result from different choices for i and s. However, in this case we wont need a precise representation of the coefficients since they will only be used to derive an O-term for the asymptotic. For $\nu = i$ we have a pole of order 2 at $s = i+1$ which implies as the dominating part of its residue

$$-\left(-2\frac{x(1+y)}{1-2x-2xy}\right)^i \frac{\ln(t)}{t}.$$

This term will provide the contribution of highest significance for $\nu = 0$, the term of second order for $\nu = 1$ and the term of third order for $\nu = 2$. For $\nu = 0$ the term of second order is implied by the residues for $i = 0$, $s = 0$ and $i = 1$, $s = 1$; again we do not need a precise representation, it is sufficient to know that those residues are of constant order. The resulting expansions of $\varsigma(e^{-t})$ around $t = 0$ can be transformed into an expansion around the dominant singularity $x_0(y) = \frac{1}{2+4y}$ because $t = 0$ corresponds to $u = 1$ and $u|_{x=\frac{1}{2+4y}} = 1$ holds. The transformation is done by replacing t by $2\frac{\sqrt{2}\sqrt{y(1+2y)}}{1+2y}\xi^{1/2}$ for $\xi := 1-(2+4y)x$. Furthermore, we have to expand $-\left(a+\frac{b}{c}\right)$ and a around $x = (2+4y)^{-1}$. The related expansions are given by

$$-\left(a+\frac{b}{c}\right) = 2\sqrt{2y(1+2y)}\xi^{1/2} + O(\xi^{3/2}) \text{ and}$$

$$a = 2y - \sqrt{2y(1+2y)}\xi^{1/2} + O(\xi).$$

Combining all the results we get the following expansions of $M_\nu(x,y)$ at $x = (2+4y)^{-1}$ (terms relevant for the asymptotic only)

$$M_\nu(x,y) = \begin{cases} -\frac{1+2y}{2}\ln(\xi) + F_1\xi^{1/2} + O(\xi^{3/2}) & : \ \nu = 0, \\ \frac{\pi^2}{6}(1+2y)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-1}\xi^{-1/2} + F_2\ln(\xi) + O(\xi^{1/2}) & : \ \nu = 1, \\ 2\zeta(3)(1+2y)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-2}\xi^{-1} + F_3\xi^{-1/2} + O(\ln(\xi)) & : \ \nu = 2, \\ (1+2y)\Gamma(\nu+1)\zeta(\nu+1)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-\nu}\xi^{-\nu/2} + F_4\xi^{-(\nu-1)/2} + O(\xi^{-(\nu-2)/2}) & : \ \nu \geq 3. \end{cases} \tag{4}$$

Here F_i, $1 \leq i \leq 4$, denotes some factors possibly depending on y for which we have not determined a precise representation. Let us assume that we have chosen the radius R of the Δ-domain such that the expansion (4) is valid for x with $|x - x_0(y)| \leq R/2$. In order to apply the O-transfer method it is not sufficient to know the expansion of a function at its dominant singularity. We also need to know that the O-term of the expansion is valid for the entire Δ-domain. We will just consider the case $\nu \geq 3$ to show how this can be concluded, the reasoning for the other cases would be exactly the same. We have to show that

$$\left| M_\nu(x,y) - (1+2y)\Gamma(\nu+1)\zeta(\nu+1)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-\nu}\xi^{-\nu/2} - F_4\xi^{-(\nu-1)/2} \right| \leq C|\xi|^{-(\nu-2)/2}$$

for some C uniform in y. Since $|\xi|^{-(\nu-2)/2}$ is bounded for x in the Δ-domain with $|x - x_0(y)| > R/2$ it is sufficient to show that the left-hand side of the inequality is bounded also. We know that $M_\nu(x,y)$ is bounded for those values of x in question. Obviously, also $(1+2y)\Gamma(\nu+1)\zeta(\nu+1)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-\nu}\xi^{-\nu/2}$ and $F_4\xi^{-(\nu-1)/2}$ are bounded for x in the Δ-domain with $|x - x_0(y)| > R/2$. Thus, the left-hand side of the inequality is bounded and we can find a C such that the above inequality is valid. Since we only consider a restricted domain for y it is obvious that C can be chosen uniformly.

Now we can apply the O-transfer method in order to approximate the coefficient $[x^\alpha]M_\nu(x,y)$ for large α. We find that $[x^\alpha]M_\nu(x,y) \sim$

$$\begin{cases} \frac{1+2y}{2}\frac{(2+4y)^\alpha}{\alpha} + \overline{F}_1\frac{(2+4y)^\alpha}{\sqrt{\alpha^3}} + O(\frac{(2+4y)^\alpha}{\sqrt{\alpha^5}}) & : \ \nu = 0, \\ \Gamma(2)\zeta(2)(1+2y)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-1}\frac{(2+4y)^\alpha}{\sqrt{\pi\alpha}} + \overline{F}_2\frac{(2+4y)^\alpha}{\alpha} + O(\frac{(2+4y)^\alpha}{\sqrt{\alpha^3}}) & : \ \nu = 1, \\ \Gamma(3)\zeta(3)(1+2y)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-2}(2+4y)^\alpha + \overline{F}_3\frac{(2+4y)^\alpha}{\sqrt{\alpha}} + O(\frac{(2+4y)^\alpha}{\alpha}) & : \ \nu = 2, \\ \Gamma(\nu+1)\zeta(\nu+1)\left(2\sqrt{\frac{2y}{1+2y}}\right)^{-\nu}\frac{(2+4y)^{\alpha+1}\alpha^{\nu/2-1}}{2\Gamma(\nu/2)} + \overline{F}_4(2+4y)^\alpha\alpha^{(\nu-3)/2} + O((2+4y)^\alpha\alpha^{(\nu-4)/2}) & : \ \nu \geq 3, \end{cases}$$

for $\alpha \to \infty$ and $\overline{F}_i$, $1 \leq i \leq 4$, some factors possibly depending on y which result form F_i and the application of the transfer. We can use the saddle point method in order to determine the coefficient at $[y^\beta]$ from the above asymptotics. In all cases we need to determine the coefficient of a function of the pattern $V(y) = A(y)B(y)^\alpha$ for large α. This can be done by the Cauchy integral

$$[y^\beta]V(y) = \frac{1}{2\pi i}\int_{|y|=y_0} V(y)y^{-\beta-1}dy,$$

where y_0 is the saddle point of the function $B(y)^\alpha y^{-\beta}$ which is given by the solution of the equation

$$\frac{y_0 B'(y_0)}{B(y_0)} = \frac{\beta}{\alpha}.$$

As a special case of the results in [6] we find that

$$[y^\beta]V(y) = \frac{A(y_0)}{\sqrt{2\pi\alpha C(y_0)}} B(y_0)^\alpha y_0^{-\beta}(1+O(\alpha^{-1})), \tag{5}$$

for $C(y) = y^2B''(y)/B(y) - y^2B'(y)^2/B(y)^2 + yB'(y)/B(y)$, $\frac{\alpha}{\beta} > 1$ and $\alpha \to \infty$. This formula can be applied to the two leading terms of each of the asymptotics for $[x^\alpha]M_\nu(x,y)$. Note that we always get a factor of the order $1/\sqrt{\alpha}$ when we apply this procedure. In order to handle the O-terms we just use saddle point bounds like those given in Theorem 6.1 of [12]. In this way we find that

$$[y^\beta]\,O((2+4y)^\alpha) = O((2+4y_0)^\alpha y_0^{-\beta}). \tag{6}$$

In such a case we do not get a factor $1/\sqrt{\alpha}$, that is why it was not sufficient to consider the leading term of the expansions only.
In all cases where we will use (5), $B(y)$ is equal to $(2+4y)$, the corresponding saddle point is given by $y_0 := 1/(2(\rho-1))$ and the resulting $C(y_0)$ is given by $(\rho-1)/\rho^2$. Recall that $\rho = \frac{\alpha}{\beta}$ was restricted to be fixed and greater than 1 such that y_0 is always well-defined.
Returning to (3) we conclude that the case $r=1$ corresponds to $a + M_0(x,y)$. Since the expansion of a possesses a term of order $\xi^{1/2}$ it provides a contribution to the term of second order of the asymptotic for the coefficient at x^α which we do not consider explicitly since it will vanish in the O-term. By the application of (5) and (6) we find that for $\frac{\alpha}{\beta} > 1$ fix the coefficient in question is given by

$$[x^\alpha y^\beta]\,(a + M_0(x,y)) = \frac{2^{\alpha+\beta-3/2}(\rho-1)^{\beta-\alpha}\rho^{\alpha+2}}{\sqrt{\pi}(\alpha(\rho-1))^{3/2}} + O(\alpha^{-2}(2+4y_0)^\alpha y_0^{-\beta}).$$

For $r \geq 2$ we observe that for each $r-j \geq 2$ fixed the contribution to the generating function for the r-th moment is given by $\binom{r}{j}M_{r-j}(x,y)$. Therefore, for $r \geq 2$ fixed, the most significant contribution is implied by the choice $j=1$, i.e., we have to set $\nu := r-1$ to use the appropriate asymptotic. Again we apply the saddle point method via equation (5) in order to get an asymptotic for the coefficient at $x^\alpha y^\beta$. This procedure yields

$$[x^\alpha y^\beta]M_{r-1}(x,y) \sim \frac{2^{\alpha+\beta-1/2}\alpha^{(r-3)/2}(r-1)(\rho-1)^{\beta-\alpha-1}\rho^{(2\alpha+r+3)/2}\Gamma(\frac{r}{2}+1)\zeta(r)}{\pi\sqrt{\alpha(\rho-1)}} + O(\alpha^{(r-5)/2}(2+4y_0)^\alpha y_0^{-\beta}),$$

for $r \geq 2$, $\rho = \frac{\alpha}{\beta} > 1$ fix and $\alpha \to \infty$. To find an asymptotic for the r-th moment this quantity must be divided by the asymptotical number of (α,β)-tries. This number can be determined in exactly the same way as done in the previous calculations. We expand $T(x,y)$ as given in (1) around the dominant singularity $x = (2+4y)^{-1}$ yielding $-\sqrt{2y}\sqrt{1+2y}\xi^{1/2}$ for the term that contributes most significantly to the asymptotic. Afterwards we apply the O-transfer method and the saddle point method to find

$$T_{\alpha,\beta} \sim \frac{2^{\alpha+\beta-3/2}(\rho-1)^{\beta-\alpha}\rho^{\alpha+3/2}}{\alpha^2\pi(\rho-1)^{3/2}}. \tag{7}$$

Remark: Note that we could also use Stirling's formula together with the representation of $T_{\alpha,\beta}$ as given in Lemma 1 in order to derive the asymptotical number of (α,β)-tries. Choosing this procedure and not approximating the factorial $(\alpha-\beta+1)!$ appearing in $\binom{\alpha-1}{\alpha-\beta+1}$, the resulting asymptotic is valid uniformly for arbitrary choices of α and β, $\alpha,\beta\to\infty$.

Now everything is prepared to conclude the main theorem of the present section. Dividing the asymptotics derived from $M_\nu(x,y)$ by the asymptotical number of (α,β)-tries provides the following theorem.

Theorem 2 *The r-th moment of the stack-size of (α,β)-tries is asymptotically given by*

$$\begin{cases} \sqrt{\pi\rho\alpha} + O(1) & : \quad r = 1, \\ \zeta(r)r(r-1)\Gamma\left(\frac{r}{2}\right)\rho^{\frac{r}{2}}\alpha^{\frac{r}{2}} + O(\alpha^{(r-1)/2}) & : \quad r \geq 2, \end{cases}$$

for $\rho = \frac{\alpha}{\beta} > 1$ *fixed,* $\alpha\to\infty$. □

Note that the limit $r \to 1$ applied to our result for $r \geq 2$ yields the leading term of the expected value.
It is possible to derive terms of lower significance from our formulæ. For example, the term of second order of the expectation is implied by the expansion of a and the residues of our Mellin transform for $i = 0$, $s = 0$ and $i = 1$, $s = 1$. Those imply the term

$$\frac{\sqrt{1+2y}(2+y)}{\sqrt{2y}}\xi^{1/2}$$

for the expansion at $x = (2+4y)^{-1}$ and thus a contribution of $\frac{3}{2} - 2\rho$ for the average value. Investigating the leading term of our asymptotic expansion more precisely proves that it does not provide any contribution to the term of second order for the asymptotics.
The class of $(\alpha,\alpha+1)$-tries is equal to the class of ordinary extended binary trees with α internal nodes. Since $\lim_{\alpha\to\infty}\frac{\alpha}{\alpha+1} = 1$ holds, those trees would correspond to the case $\rho = 1$. However, even if we are formally not allowed to do so, it is sufficient to set $\rho := 1$ within the asymptotic of the expectation in order to rediscover the result of [2] (leading and constant term); there it has been shown that the average stack-size of extended binary trees is asymptotically given by $\sqrt{\pi\alpha} - \frac{1}{2}$. Furthermore, we can use our representation for the moments to determine an asymptotic for the variance. We find:

Corollary 1 *The variance $\sigma^2(\alpha,\beta)$ of the stack-size of (α,β)-tries is asymptotically given by*

$$\sigma^2(\alpha,\beta) \sim \left(\frac{1}{3}\pi\rho - \rho\right)\pi\alpha,$$

for $\rho = \frac{\alpha}{\beta} > 1$ *fixed,* $\alpha\to\infty$. □

Note, that again $\rho = 1$ leads to the well-known result for ordinary extended binary trees (see e.g. [15]).
Let the *normalized stack-size* of a (α,β)-trie T be defined as $\hat{s}(T) := s(T)/\sqrt{\alpha\rho}$ for $\rho = \frac{\alpha}{\beta}$. Then we obviously find that the r-th moment of the normalized stack-size is asymptotically given by

$$r(r-1)\Gamma\left(\frac{r}{2}\right)\zeta(r)$$

in the limit. Those are exactly the r-th moments of the theta distribution [27] whose cummulative distribution function is

$$H(x) = 4x^{-3}\pi^{\frac{5}{2}} \sum_{k\geq 0} k^2 \exp(-k^2\pi^2/x^2)$$

with the corresponding density

$$h(x) = 4x \sum_{k\geq 1} k^2(2k^2x^2 - 3)\exp(-k^2x^2). \tag{8}$$

Therefore we can conclude:

Corollary 2 *The normalized stack-size of* (α,β)*-tries*

$$\hat{s}(T) = s(T)/\sqrt{\alpha\rho}$$

admits a limiting theta distribution with density function (8) for $\rho = \frac{\alpha}{\beta} > 1$ *fixed,* $\alpha \to \infty$*.* □

See [15] for the related result for ordinary extended binary trees which we *rediscover* by setting ρ to 1. As already mentioned within the introduction there are also studies for combinatorial tries where only the number of internal nodes is used to determine the size of a tree. In the sequel we will call a $\mathcal{C}$-trie with α internal nodes and an arbitrary number of white leaves a α-trie. Since there are no results on higher moments, distribution or variance for the stack-size of uniform random α-tries, we will use our computations in order to derive the related approximations. Setting $y = 1$ within the asymptotic for $[x^\alpha]M_{r-1}(x,y)$ yields the corresponding asymptotic for α-tries

$$3r\Gamma(r)\zeta(r)\left(\frac{2}{3}\sqrt{6}\right)^{-(r-1)} 6^\alpha \alpha^{\frac{r}{2}-\frac{3}{2}}\Gamma^{-1}\left(\frac{r-1}{2}\right).$$

Dividing this quantity by the asymptotical number of α-tries

$$\frac{\sqrt{6}6^\alpha}{2\sqrt{\pi\alpha^3}}$$

leads to the approximation of the r-th moment given in the following corollary.

Corollary 3 *For* $r \geq 2$ *fixed, the r-th moment of the stack-size of* α*-tries is asymptotically given by*

$$\zeta(r)(r-1)r\left(\frac{3}{2}\right)^{\frac{r}{2}} \alpha^{\frac{r}{2}}\Gamma\left(\frac{r}{2}\right),$$

$\alpha \to \infty$. □

Note, that even if we only considered $r \geq 2$ for the α-tries, the limit $r \to 1$ is equal to the well-known approximation for the average-value $\sqrt{\frac{3}{2}\pi\alpha}$ which was first proven in [23]. Furthermore, it is obvious to normalize the stack-size of α-tries in order to find again a limiting theta-distribution. Moreover, we can use our results to prove an asymptotic for the variance. We find:

Corollary 4 *The variance $\sigma^2(\alpha)$ of the stack-size of α-tries is asymptotically given by*

$$\frac{1}{2}\pi(\pi-3)\alpha, \ \alpha \to \infty.$$

If the normalized *stack-size $\tilde{s}(T)$ of a α-trie T is defined as $\tilde{s}(T) := s(T) / \left(\frac{\sqrt{6}}{2}\sqrt{\alpha}\right)$, then $\tilde{s}(T)$ admits a limiting theta distribution with density function (8) for $\alpha \to \infty$.*

□

We will finish this section by providing some plots and tables related to the results presented. In Figure 5

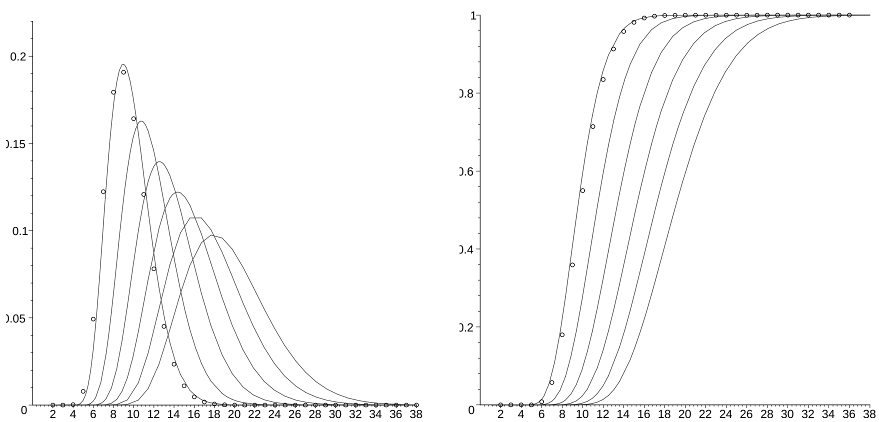

Fig. 5: The density and the distribution function for (α,β)-tries for $\beta = 30$ and $\rho = \frac{i}{5}$, $i = 5, 6, \ldots, 10$. The solid lines represent the asymptotics, the circles are used to plot the exact values.

we find the asymptotical density and distribution function for the stack-size of (α,β)-tries for $\beta = 30$ and ρ restricted to some fixed values compared to the exact values of the density and distribution. For the density (left picture of Figure 5) the highest graph corresponds to the case $\rho = 1$ and the lowest one to the case $\rho = 2$. For the distribution function the leftmost graph belongs to $\rho = 1$ and the rightmost to $\rho = 2$. In both pictures the asymptotic is only compared with the exact values in the case $\rho = 1$. As you can see our predictions are very close to the real values even for $\mathcal{C}$-tries of relatively small sizes. In Figure 6 similar plots are pictured for the case of the normalized stack-size for α-tries of size 50. Again, we find that the presented asymptotics are very accurate. Finally, the table of Figure 7 shows some exact numerical values of the distribution function for the normalized stack-size of α-tries together with the corresponding approximations.

4 Concluding Remarks

In this paper we have investigated the average stack-size of uniform random $\mathcal{C}$-tries with α internal nodes and β white leaves (keys). Our result improves the one presented in [23] because the number of white leaves (code words, keys) within a $\mathcal{C}$-trie has been introduced as a new parameter. With respect to the application of tries to the compression of blockcodes the ratio $\rho := \frac{\alpha}{\beta}$ can be considered as the degree of

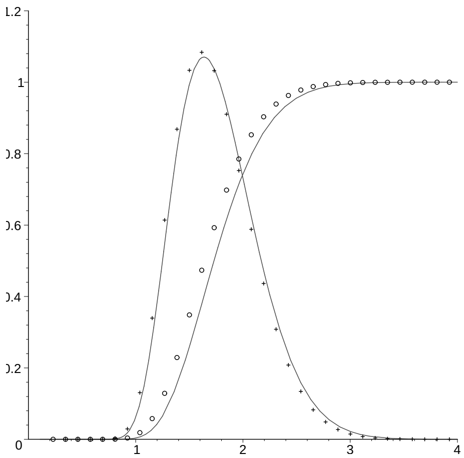

Fig. 6: The density and the distribution function for α-tries of size 50. The solid lines represent the asymptotics, the circles and crosses are used to plot the exact values.

$\Pr[\tilde{s}(T) \le p]$				
	p			
α	1	2	3	4
6	0.213	0.981	1.000	1.000
24	0.073	0.907	0.999	0.999
54	0.038	0.867	0.999	0.999
96	0.025	0.842	0.998	0.999
150	0.019	0.825	0.998	0.999
216	0.015	0.814	0.997	0.999
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	0.003	0.743	0.995	0.999

Fig. 7: Some values of the distibution function of the normalized stack-size. The last row represents the asymptotic values for $\alpha \to \infty$.

redundancy of the compressed code. Thus, the introduction of the new parameter makes it possible to quantify the additional costs, which are implied by a certain amount of redundancy. Furthermore, it was possible to determine the distribution (in the limit) of the stack-size for the class of $\mathcal{C}$-tries. All the results for (α, β)-tries presented in this paper can be considered as a generalization of related results for ordinary extended binary trees published in [2] and [15]. We rediscover those results by setting the ratio $\frac{\alpha}{\beta}$ to 1 within our formulæ. A similar parameter, the so-called Horton-Strahler number, has been considered in [24]. In this paper the author has proved that (under the same assumptions as in Theorem 2 of the present paper) the average Horton-Strahler number of large $\mathcal{C}$-tries is independent of the number of white leaves. This supports the conjecture stated in [23] that there is no simple relation between extended binary trees and $\mathcal{C}$-tries (colored extended binary trees), which would allow us to conclude our results from the well known related results for extended binary trees; such a relation should work in the same manner for both parameters.

There is a recent work by J. Bourdon, B. Vallée and the author [3] in which we investigated the stack-size of m-ary tries under probability models which are more natural for tries considered as an efficient implementation of a dictionary. We have shown in [3], that the stack-size of a trie built from n keys independently emitted by a source has an expectation of order $\log n$ and a probability distribution which is asymptotically of the double exponential type.

Acknowledgements: I wish to thank Michael Drmota for some useful hints concerning multivariate asymptotics and also the anonymous referee, whose suggestions helped me to improve the quality of the paper.

References

[1] T. M. APOSTOL, *Introduction to Analytic Number Theory*, Springer, 1976.

[2] N. G. DE BRUIJN, D. E. KNUTH, S.O. RICE, *The Average Height of Planted Plane Trees*, Graph Theory and Computing (R. C. Read, ed.), Academic Press, 1972.

[3] J. BOURDON, M. NEBEL, B. VALLÉE, *On the Stack-Size of General Tries*, R.A.I.R.O. Theoretical Informatics and Applications **35**, 2001, 163-185.

[4] J. CLÉMENT, P. FLAJOLET AND B. VALLÉE, *Dynamical Sources in Information Theory: A General Analysis of Trie Structures*, Algorithmica **29**, 2001, 307-369.

[5] L. DEVROYE, *A probabilistic analysis of the height of tries and the complexity of triesort*, Acta Informatica **21**, 1984, 229-237.

[6] M. DRMOTA, *A Bivariate Asymptotic Expansion of Coefficients of Powers of Generating Functions*, European Journal of Combinatorics **15**, 1994, 139-152.

[7] P. FLAJOLET, *On the Performance Evaluation of Extendible Hashing and Trie Searching*, Acta Informatica **20**, 1983, 345-369.

[8] P. FLAJOLET, X. GOURDON AND P. DUMAS, *Mellin transforms and asymptotics: Harmonic sums*, Theoretical Computer Science **144**, 1995, 3-58.

[9] P. FLAJOLET AND A. ODLYZKO, *The Average Height of Binary Trees and Other Simple Trees*, Journal of Computer and System Sciences **25**, 1982, 171-213.

[10] P. FLAJOLET AND A. ODLYZKO, *Singularity Analysis of Generating Functions*, SIAM J. Disc. Math. **3**, No. 2, 1990, 216-240.

[11] P. FLAJOLET AND R. SEDGEWICK, *Analysis of Algorithms*, Addison-Wesley Publishing Company, 1996.

[12] P. FLAJOLET AND R. SEDGEWICK, *The average case analysis of algorithms: Saddle Point Asymptotics*, INRIA Rapport de recherche **2376**, 1994.

[13] G. H. GONNET AND R. BAEZA-YATES, *Handbook of Algorithms and Data Structures*, Addison-Wesley, 1991.

[14] P. JACQUET AND M. RÉGNIER, *Trie Partitioning Process: Limiting Distributions*, Lecture Notes in Computer Science **214**, 1986, 196-210.

[15] R. KEMP, *On the Average Stack-Size of Regularly Distributed Binary Trees*, Proc. ICALP 79, Lecture Notes in Computer Science **71**, 1979, 340-355.

[16] R. KEMP, *The Average Height of r-Tuply Rooted Planted Plane Trees*, Computing **25**, 1980, 209-232.

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

[18] P. KIRSCHENHOFER, H. PRODINGER AND W. SZPANKOWSKI, *On the Variance of the External Path Length in a Symmetric Digital Trie*, Discrete Applied Mathematics **25**, 1989, 129-143.

[19] P. KIRSCHENHOFER, H. PRODINGER AND W. SZPANKOWSKI, *On the Balance Property of PATRICIA Tries: External Path Length Viewpoint*, Theoretical Computer Science **68**, 1989, 1-17.

[20] D. E. KNUTH, *The Art of Computer Programming*,
Volume 3: Sorting and Searching, Addison-Wesley, 1973.

[21] J. R. LIEBEHENSCHEL, *Ranking and Unranking of Lexicographically Ordered Words: An Average-Case Analysis*, Journal of Automata, Languages and Combinatorics **2** (4), 1997, 227-268.

[22] H. M. MAHMOUD, *Evolution of Random Search Trees*, Wiley-Interscience Series, 1992.

[23] M. E. NEBEL, *The Stack-Size of Tries, A Combinatorial Study*, Theoretical Computer Science **270**, 2002, 441-461.

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

[25] B. RAIS, P. JACQUET AND W. SZPANKOWSKI, *A Limiting Distribution for the Depth in Patricia Tries*, SIAM J. Discrete Mathematics **6**, 1993, 197-213.

[26] M. RÉGNIER AND P. JACQUET, *New Results on the Size of Tries*, IEEE Transactions on Information Theory **35**, 1989, 203-205.

[27] A. RENYI AND G. SZEKERES, *On the height of trees*, Austral. J. Math. **7**, 1967, 497-507.

[28] R. SEDGEWICK, *Algorithms*, Addison-Wesley, 1988.

[29] W. SZPANKOWSKI, *Some Results on V-ary Asymmetric Tries*, Journal of Algorithms **9**, 1988, 224-244.

[30] W. SZPANKOWSKI, *On the Height of Digital Trees and Related Problems*, Algorithmica **6**, 1991, 256-277.

[31] W. SZPANKOWSKI, *Patricia Tries Again Revisited*, J. ACM **37**, 1990, 691-711.