DISTRIBUTION OF VARIABLES IN LAMBDA-TERMS WITH RESTRICTIONS ON DE BRUIJN INDICES AND DE BRUIJN LEVELS

BERNHARD GITTENBERGER AND ISABELLA LARCHER

ABSTRACT. We investigate the number of variables in two special subclasses of lambda-terms that are restricted by a bound of the number of abstractions between a variable and its binding lambda, the so-called De-Bruijn index, or by a bound of the nesting levels of abstractions, *i.e.*, the number of De Bruijn levels, respectively. These restrictions are on the one hand very natural from a practical point of view, and on the other hand they simplify the counting problem compared to that of unrestricted lambda-terms in such a way that the common methods of analytic combinatorics are applicable.

We will show that the total number of variables is asymptotically normally distributed for both subclasses of lambda-terms with mean and variance asymptotically equal to Cn and $\tilde{C}n$, respectively, where the constants C and $\tilde{C}$ depend on the bound that has been imposed. So far we just derived closed formulas for the constants in case of the class of lambda-terms with bounded De Bruijn index. However, for the other class of lambda-terms that we consider, namely lambda-terms with a bounded number of De Bruijn levels, we investigate the number of variables, as well as abstractions and applications, in the different De Bruijn levels and thereby exhibit a so-called "unary profile" that attains a very interesting shape.

1. INTRODUCTION

1.1. **Motivation.** Lambda-terms are objects stemming from lambda-calculus and can be seen as combinatorial objects with a simple description. Nevertheless, the enumeration of lambda-terms is not well understood. Combinatorially, they can be seen as words (sequence of symbols) or graphs and thus the combinatorially most natural way to define an enumeration problem is to ask for the number of terms with a given number of symbols or vertices, respectively. This problem appears to be very intriguing, as standard techniques fail. Considering subclasses by imposing certain restrictions can turn the enumeration problem into an accessible one, but for one particular model the enumeration formulas exhibit a very peculiar behaviour. Our motivation to perform the present investigation is to shed light on this oddity and to give a combinatorial explanation for this phenomenon.

1.2. **Previous work and the considered model.** Lambda-calculus is a set of rules to manipulate lambda-terms and it is an important tool in theoretical computer science. To our knowledge, the first appearance of enumeration problems in the sense of enumerative combinatorics which are linked to lambda-calculus is found in [25], where certain models of lambda-calculus are analyzed which have representations as formal power series. More recently, we observe rising interest in the quantitative properties of large random lambda-terms. The first work in this direction seems to be [28]. Later David *et al.* [15] investigated the proportion of normalising terms, which was also the topic of [5] in a different context. Other papers dealing with certain structural properties of lambda-terms are for instance [13, 22, 30].

Since studying quantitative aspects of lambda-terms using combinatorial methods relies heavily on their enumeration, many papers are devoted to their enumeration, which itself very much depends on the particular class of terms and the definition of the term size. The enumeration may be done by constructing bijections to certain classes of maps, see e.g. [6, 34, 35] or the use of the methodology from analytic combinatorics [20], see e.g. [3, 7, 8, 9, 11, 24, 27].

Another approach to gain structural insight is by random generation. Solving the enumeration problems is the basis for an efficient algorithm for this purpose, namely Boltzmann sampling [17, 18]. The method is extendible to a multivariate setting allowing for a fine tuning according to specified structural properties of the sampled objects, as was demonstrated in [2, 12]. The generation of lambda-terms was treated in [4, 6, 24, 29, 31, 32].

As mentioned above, lambda-terms can be seen as graphs, and so the natural combinatorial question is to ask for the number of such graphs with n vertices. It turned out, however, that this model leads to a generating function being a purely formal power series, which can be represented as an infinite nesting of square-roots of polynomial terms, see [7]. The enumeration problem becomes amenable to analytic techniques, when imposing restrictions on the lambda-terms: One such model is the class of all lambda-terms with a bounded number of De Bruijn levels. In [8] the authors discovered a very interesting phenomenon concerning the generating function associated with this model, namely that the asymptotic behaviour of the coefficients of the generating function changes with the imposed bound. More precisely, the type of the dominant singularity changes from $\frac{1}{2}$ to $\frac{1}{4}$ whenever the bound belongs to a special doubly-exponentially growing sequence. This alteration of the type of the dominant singularity has a direct impact on the subexponential factor of the coefficient's asymptotics, namely it shifts from $n^{-3/2}$ to $n^{-5/2}$ for n tending to infinity. This paper studies the structure of random large lambda-terms belonging to this class and thereby delivers an explanation of the above mentioned phenomenon, since it arises from the location of the variables within the lambda-term.

The lambda-calculus was invented by Church and Kleene in the 1930ies as a tool for the investigation of decision problems. Today it still plays an important role in computability theory and for automatic proof systems. Furthermore, it represents the basis for some programming languages, such as LISP or Haskell. In fact, the generation of random lambda terms served for optimising the Glasgow Haskell Compiler [29] and for finding bug in a C-compiler [33] As mentioned at the beginning, recently, rising interest in the number and structural properties of lambda-terms can be observed. This is triggered on the one hand by the fact that random lambda-terms have practical application and the understanding of structural properties enables their tuning when generating random terms, see [2], on the other hand, they turned out to be a source of interesting, albeit in part very intricate, combinatorial enumeration problems. Finally, we mention that there is a direct relationship between these random structures acting as computer programs and mathematical proofs (see [14]), but this relationship essentially concerns only typed lambda-terms and not general ones.

For a thorough introduction to lambda calculus we refer to [1]. This paper does not require any preliminary knowledge of lambda calculus in order to follow the proofs. Instead we will study the basic objects of lambda calculus, namely lambda-terms, by considering them as combinatorial objects, or more precisely as a special class of directed acyclic graphs (DAGs).

1.3. **The considered model.** We are now in position to formally define the model of lambda-terms which we are considering in this paper. The way we see the structures and the methodology of proving our results heavily relies on analytic combinatorics. Thus, a profound knowledge of some methods in this field is necessary to follow all our arguments in detail. The prerequisites can be found in Flajolet and Sedgewick's seminal book [20].

Definition 1 (lambda-terms, [23, Definition 3]). Let $\mathcal{V}$ be a countable set of variables. The set Λ of lambda-terms is defined by the following grammar:

(1) every variable in $\mathcal{V}$ is a lambda-term,
(2) if T and S are lambda-terms then TS is a lambda-term, (application)
(3) if T is a lambda-term and x is a variable then $\lambda x.T$ is a lambda-term. (abstraction)

The name application arises, since lambda-terms of the form TS can be regarded as functions $T(S)$, where the function T is applied to S, which in turn can be a function itself. An abstraction can be considered as a quantifier that binds the respective variable in the sub-lambda-term within its scope. Both application and repeated abstraction are not commutative, *i.e.*, in general the

lambda-terms TS and ST, as well as $\lambda x.\lambda y.M$ and $\lambda y.\lambda x.M$, are different (with the exceptions of $T = S$ and none of the variables x or y occurring in M, respectively). Each λ binds exactly one variable, which may occur several times in a term or even not at all. And since we will just focus on the special subclass of closed lambda-terms, each variable is bound.

We will consider lambda-terms modulo α-equivalence, which means that we identify two lambda-terms if they only differ by the names of their bound variables. For example $\lambda x.(\lambda y.(xy)) \equiv \lambda y.(\lambda z.(yz))$. 1972 De Bruijn [16] introduced a representation for lambda-terms that completely avoids the use of variables by substituting them by natural numbers that indicate the number of abstractions between the variable and its binding lambda (the binding lambda is counted as well), i.e., $\lambda x.(\lambda y.(xy)) = \lambda(\lambda 21)$.

Definition 2 (De Bruijn index, De Bruijn level). The natural numbers that represent the variables in the De Bruijn representation of a lambda-term are called De Bruijn indices. The number of nested lambdas starting from the outermost one specifies the De Bruijn level in which a variable (or De Bruijn index, respectively) is located.

For example in the lambda-term $\lambda x.x(\lambda y.(xy)) = \lambda 1(\lambda 21)$ the first occurrence of the variable x (i.e., the leftmost 1 in the De Bruijn representation) is in the first De Bruijn level, while the other variables are in the second De Bruijn level.

There is also a combinatorial interpretation of lambda-terms that considers them as DAGs and thereby naturally identifies two α-equivalent terms to be equal. Combinatorially, lambda-terms can be seen as rooted unary-binary trees containing additional directed edges. Note that in general the resulting structures are not trees in the sense of graph theory, but due to their close relation to trees (see Definition 3) some authors call them lambda-trees or enriched trees. We will call them lambda-DAGs in order to emphasise that these structures are in fact DAGs, if we consider the undirected edges of the underlying tree to be directed away from its root.

Definition 3 (lambda-DAG, [23, Definition 5]). With every lambda-term T, the corresponding lambda-DAG $G(T)$ can be constructed in the following way:

(1) If x is a variable then $G(x)$ is a single node labeled with x. Note that x is unbound.
(2) $G(PQ)$ is a lambda-DAG with a binary node as root, having the two lambda-DAGs $G(P)$ (to the left) and $G(Q)$ (to the right) as subgraphs.
(3) The DAG $G(\lambda x.P)$ is obtained from $G(P)$ in four steps:
   (a) Add a unary node as new root.
   (b) Connect the new root by an undirected edge with the root of G(P).
   (c) Connect all leaves of $G(P)$ labelled with x by directed edges with the new root, where the root is start vertex of these edges.
   (d) Remove all labels x from $G(P)$. Note that now x is bound.

Obviously, applications correspond to binary nodes and abstractions correspond to unary nodes of the underlying Motzkin-tree that is obtained by removing all directed edges. Of course, in the lambda-DAG some of the vertices that were former unary nodes might have gained out-going edges, so they are no unary nodes in the lambda-DAG anymore. However, when we speak of unary nodes in the following, we mean the unary nodes of the underlying unary-binary tree that forms the skeleton of the lambda-DAG.

Since the skeleton of a lambda-DAG is a tree, we sometimes call the variables leaves (i.e., the nodes with out-degree zero), and the path connecting the root with a leaf (consisting of undirected edges) is called a branch. There are different approaches as to how one can define the size of a lambda-term (see [15, 8, 27]), but within this paper the size will be defined as the number of nodes in the corresponding lambda-DAG. This is combinatorially the most natural definition, and it is even equivalent to Barendregt's [1] definition.

In the lambda-DAG, the De Bruijn indices and levels are easily visible: The De Bruijn index of a variable v is the number of unary nodes we find on the path from v to its binding lambda in the skeleton of the lambda-DAG, where the last unary node on the path has to be counted as well. The De Bruijn level of v is the number of unary nodes on the path from v to the root. See Figure 2.

(A) $\lambda x.((\lambda y.(xy))x)$
$= \lambda((\lambda 21)1)$

(B) $(\lambda x.(x(\lambda y.y))(\lambda x.(\lambda y.xy))$
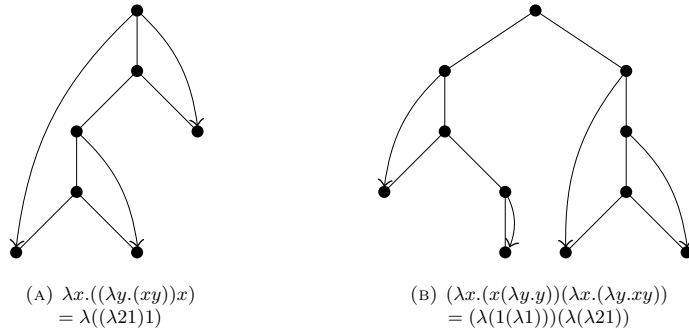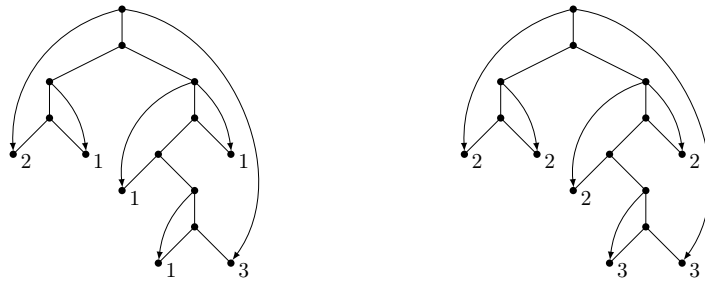$= (\lambda(1(\lambda 1)))(\lambda(\lambda 21))$

FIGURE 1. The lambda-DAGs corresponding to the respective terms written below.

FIGURE 2. The lambda-DAG representing the term $\lambda x.((\lambda y.xy)(\lambda z.(z(\lambda t.tx))z))$, where the leaves are labelled with (left) the corresponding De Bruijn indices, and (right) the De Bruijn level in which they are located.

At first sight lambda-terms appear to be very simple structures, in the sense that their construction can easily be described, but so far no one has yet accomplished to derive their asymptotic number. However, the asymptotic equivalent of the logarithm of this number can be determined up to the second-order term (see [9]). The difficulty of counting unrestricted lambda-terms arises due to the fact that their number increases superexponentially with increasing size. The reason for this are the many degrees of freedom to choose the bindings. Thus, if we translate the counting problem into generating functions, then the resulting generating function has a radius of convergence equal to zero, which makes the common methods of analytic combinatorics inapplicable. This fast growth of the number of lambda-terms can be explained by the numerous possible bindings of leaves by lambdas, *i.e.* by unary nodes. Consequently, lately some simpler subclasses of lambda-terms, which reduce these multiple binding possibilities, have been studied, e.g. lambda-terms with prescribed number of unary nodes ([8]), or lambda-terms in which every lambda binds a prescribed ([9, 6, 23]) or a bounded ([10, 6, 23]) number of leaves. In this paper we will investigate structural properties of lambda-terms that have been introduced in [7] and [8], namely at first lambda-terms with a bounded number of abstractions between each leaf and its binding lambda, which corresponds to a bounded De Bruijn index. The second class of lambda-terms that we will investigate within this paper is the class of lambda-terms with a bounded number of nesting levels of abstractions, *i.e.*, lambda-terms with a bounded number of De Bruijn levels. From a practical point of view these restrictions appear to be very natural, since the number of abstractions in lambda-terms which are used for computer programming is in general assumed to be very low compared to their size (*cf.* [33]).

Particular interest lies in the number and distribution of the variables within these special subclasses of lambda-terms. We will show within this paper that the total number of leaves (*i.e.*, variables) in lambda-DAGs with bounded De Bruijn indices as well as in lambda-terms with bounded number of De Bruijn levels is asymptotically normally distributed. For the latter class of lambda-terms we will also investigate the number of leaves in the different De Bruijn levels,

which shows a very interesting behaviour. We will see that in the lower De Bruijn levels, *i.e.* near the root of the lambda-DAG, there are very few leaves, while almost all of the leaves are located in the upper De Bruijn levels and these two domains will turn out to be asymptotically strictly separated. The same behaviour can be shown for unary and binary nodes, which allows us to set up a very interesting "unary profile" of this class of lambda-terms.

For lambda-terms that are locally restricted by a bound for the De Bruijn indices the number of De Bruijn levels is not bounded and will tend to infinity for increasing size. The expected number of De Bruijn levels is unknown, which implies that the correct scaling cannot be determined. Thus, we have not been able to establish results concerning the leaves (or other types of nodes) on the different De Bruijn levels for this class of lambda-terms so far. Nevertheless, further studies on this subject seem to be very interesting.

Plan of the paper. We will present the main results that have been derived in this paper, including all the definitions that are necessary for their understanding, in Section 2, while the subsequent sections are concerned with their proofs. In Section 3 we will show that the total number of variables in lambda-terms with bounded De Bruijn index is asymptotically normally distributed with mean and variance asymptotically Cn and $\tilde{C}n$, respectively, where the constants C and $\tilde{C}$ depend on the bound that has been imposed. Section 4 shows the same result for lambda-terms where the number of De Bruijn levels is bounded. Finally, in the last section, Section 5, we show how the variables are distributed in lambda-terms with bounded number of De Bruijn levels. We will see that there are very few leaves on the lower De Bruijn levels, *i.e.*, close to the root, while on the upper De Bruijn levels farther away from the root, there are many leaves. Furthermore, these two domains are strictly separated and we know exactly which is the first level containing a large number of leaves, since this level can be determined by the imposed bound of the number of De Bruijn levels. This interesting behaviour also holds for the number of binary and unary nodes. By investigating all these numbers among the different De Bruijn levels we are able to set up a so-called unary profile that shows that these special lambda-terms have a very specific shape. A random closed lambda-term with a bounded number of De Bruijn levels starts with a string of unary nodes, where the length of this string depends on the imposed bound. Then it gets slowly filled with nodes until it reaches the aforementioned separating level, where it suddenly starts to contain a lot of nodes.

2. Main results

In this section we will introduce the basic definitions and summarize the main results that will be presented in this paper.

First, we will investigate the total number of variables in lambda-terms with bounded De Bruijn index, *i.e.*, with a bounded number of abstractions between each leaf and its binding lambda. Our first main result concerns the asymptotic distribution of the number of variables within this class of closed lambda-terms.

Theorem 1. *Let X_n be the total number of variables in a random closed lambda-term of size n where the De Bruijn index of each variable is at most k. Then X_n is asymptotically normally distributed with*

$$\mathbb{E}X_n \sim \frac{k}{\sqrt{k}+2k}n, \quad and \quad \mathbb{V}X_n \sim \frac{k^2}{2\sqrt{k}(\sqrt{k}+2k)^2}n, \quad as\ n \to \infty.$$

Remark. Note that $\mathbb{E}X_n \longrightarrow \frac{n}{2}$ and $\mathbb{V}X_n \longrightarrow 0$ for $k \to \infty$. Since these values are known for the number of leaves in binary trees, this gives a hint that almost all leaves of a large random unrestricted lambda-term are located within an almost purely binary structure.

Next we turn to lambda-terms with a bounded number of De Bruijn levels, *i.e.* with a bounded number of unary nodes (or abstractions, respectively) in the separate branches of the corresponding lambda-DAG. Figure 3 shows the partition of the vertex set which is induced by the De Bruijn levels.

Remark. De Bruijn [16] introduced the concepts De Bruijn index and De Bruijn level exclusively for variables and called them reference depth and level, respectively. The notion De Bruijn level, however, can evidently be extended to all vertices of the lambda-DAG. We will use this extension later in the paper.
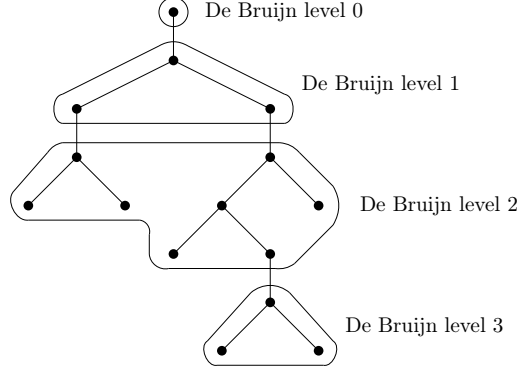
(*Cf.* also Figure 3)

FIGURE 3. As an example, we present the underlying Motzkin tree of the lambda term $\lambda x.((\lambda y.yx)(\lambda z.(z(\lambda t.tx))z))$, where the different De Bruijn levels are encircled.

Lambda-terms with bounded number of De Bruijn levels have been studied in [8], where a very unusual behaviour has been discovered. The asymptotic behaviour of the number of lambda-terms belonging to this subclass differs depending on whether the imposed bound is an element of a certain sequence $(N_i)_{i\geq 0}$, which will be given in Definition 4, or not. Though the behaviour of the counting sequence differs for these two cases, the result in Theorem 2 concerning lambda-terms with bounded number of De Bruijn levels is the same after all. However, the method of proof is different in the two cases. For our subsequent results the distinction of cases will have an impact on the asymptotic behaviour of the counting sequence of the investigated structures. Thus, we will have to distinguish between these two cases.

Definition 4 (auxiliary sequences $(u_i)_{i\geq 0}$ and $(N_i)_{i\geq 0}$, [8, Def.6])**.** Let $(u_i)_{i\geq 0}$ be the integer sequence defined by
$$u_0 = 0, \quad u_{i+1} = u_i^2 + i + 1 \quad \text{for } i \geq 0,$$
and $(N_i)_{i\geq 0}$ by
$$N_i = u_i^2 - u_i + i, \quad \text{for all } i \geq 0.$$

In the last section we investigate the distribution of the different types of nodes in lambda-DAGs with bounded number of De Bruijn levels among the separate levels throughout the DAG.

Remark. Note that the De Bruijn level in which a node is located just counts the number of unary nodes in the branch connecting the root and the respective node.

Theorem 2. *Let $(N_i)_{i\geq 0}$ be the sequence defined in Definition 4 and choose integers k and j such that $N_{j-1} < k \leq N_j$. Moreover, let $\rho_k(u)$ be the root of smallest modulus of the function $z \mapsto R_{j+1,k}(z,u)$, where*

$$R_{j+1,k}(z,u) = 1 - 4(k-j)z^2u - 2z + 2z\sqrt{1 - 4(k-j+1)z^2u - 2z + \sqrt{... + 2z\sqrt{1 - 4kz^2u}}},$$

and let us define $B(u) = \frac{\rho_k(1)}{\rho_k(u)}$.

If $B''(1) + B'(1) - B'(1)^2 \neq 0$, then the total number of leaves in random closed lambda-DAGs with at most k De Bruijn levels is asymptotically normally distributed with asymptotic mean μn and asymptotic variance $\sigma^2 n$, where $\mu = B'(1)$ and $\sigma^2 = B''(1) + B'(1) - B'(1)^2$.

Remark. The requirement $B''(1)+B'(1)-B'(1)^2 \neq 0$ obviously results from the fact that otherwise the variance would be $o(n)$ (*cf.* the variability condition in Hwang's Quasi-Powers Theorem [26]). However, this inequality seems to be very difficult to verify, since $B(u) = \frac{\rho_k(1)}{\rho_k(u)}$ and we do not know anything about the function $\rho_k(u)$, except for some crude bounds and its analyticity. But numerical data supports the conjecture that $B''(1) + B'(1) - B'(1)^2 \neq 0$ always holds (*cf.* Table 1).

The following theorem includes the results that we will present in Section 5.1, where we show that the number of leaves near the root of the lambda-DAG, *i.e.*, in the lower De Bruijn levels, is very low, while there are many leaves in the upper levels. Furthermore these two domains are strictly separated and the "separating level", *i.e.*, the first level with many leaves, depends on the bound of the number of De Bruijn levels. We will show a very interesting behaviour, namely that with growing bound the number of leaves within the De Bruijn level that is directly below the critical separating level increases, until the bound reaches a certain number, which makes this adjacent leaf-filled level become the new separating level. Thus, we can observe a "double jump" in the asymptotic behaviour of the number of leaves within the separate levels (*cf.* Figure 4).

Theorem 3. *Let $_{k-l}\tilde{H}_k(z,u)$ denote the bivariate generating function of the class of closed lambda-terms with at most k De Bruijn levels, where z marks the size and u marks the number of leaves in the $(k-l)$-th De Bruijn level. Additionally, we denote its dominant singularity (i.e., the singularity of smallest modulus) by $\tilde{\rho}_k(u)$, and $\tilde{B}(u) = \frac{\tilde{\rho}_k(u)}{\tilde{\rho}_k(1)}$. Then the following assertions hold:*

- *If $k \in (N_j, N_{j+1})$, then the average number of leaves in the first $k - j$ De Bruijn levels is $O(1)$, as the size $n \to \infty$, while it is $\Theta(n)$ for the last $j + 1$ levels.*

    In particular, if $\tilde{B}''(1) + \tilde{B}'(1) - \tilde{B}'(1)^2 \neq 0$, the number of leaves in each of the last $j + 1$ De Bruijn levels is asymptotically normally distributed with mean and variance proportional to the size n of the lambda-term.

- *If $k = N_j$, then the average number of leaves in the first $k - j$ De Bruijn levels is $O(1)$, as $n \to \infty$, while the average number of leaves in the $(k - j)$-th level is $\Theta(\sqrt{n})$. The last j De Bruijn levels have asymptotically $\Theta(n)$ leaves.*

    In particular, if $\tilde{B}''(1) + \tilde{B}'(1) - \tilde{B}'(1)^2 \neq 0$, the number of leaves in each of the last j De Bruijn levels is asymptotically normally distributed with mean and variance proportional to the size n of the lambda-term..
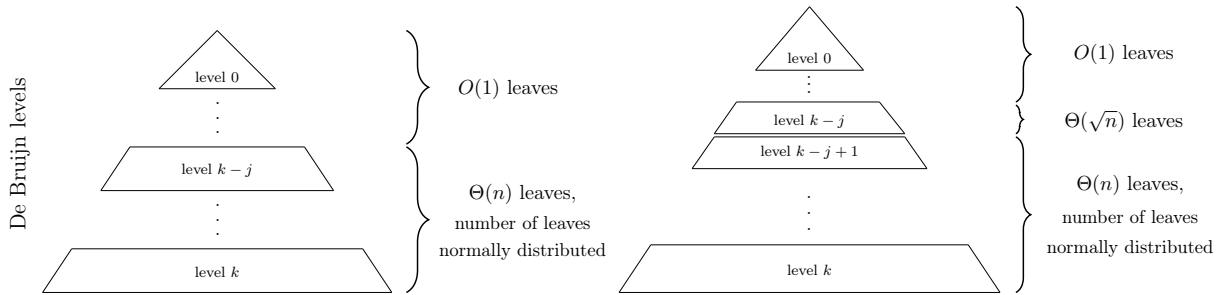

FIGURE 4. Summary of the mean values of the number of leaves in the different De Bruijn levels in lambda-terms with at most k De Bruijn levels for the case $N_j < k < N_{j+1}$ (left), and the case $k = N_j$ (right).

Sections 5.2 and 5.3 are concerned with the investigation of the number of unary nodes, and binary nodes respectively, among the De Bruijn levels. Using the same techniques as in Section 5.1 we can show that their number behaves in fact very similar to the number of leaves.

Theorem 4. *If $k \in (N_j, N_{j+1})$, then both the average number of unary nodes and the average number of binary nodes in the first $k - j$ De Bruijn levels are $O(1)$, as $n \to \infty$, while they are $\Theta(n)$ in each of the last $j + 1$ levels.*

If $k = N_j$, then both the average number of unary nodes and the average number of binary nodes in the first $k - j$ De Bruijn levels is $O(1)$, as $n \to \infty$, while the average number of nodes of the respective type in the $(k-j)$-th De Bruijn level is $\Theta(\sqrt{n})$. The last j De Bruijn levels contain each asymptotically $\Theta(n)$ unary nodes, as well as $\Theta(n)$ binary nodes.

3. Total number of leaves in random lambda-terms with bounded De Bruijn indices

In this section we investigate the asymptotic number of all leaves in closed lambda-terms with bounded De Bruijn indices. In order to get some quantitative results concerning this restricted class of lambda-terms we will use the well-known symbolic method in conjunction with analytic combinatorics (see [20]), in particular singularity analysis of generating functions [19]. Therefore we introduce further combinatorial classes as it has been done in [8]: $\mathcal{Z}$ denotes the class of atoms, $\mathcal{A}$ the class of application nodes (*i.e.*, binary nodes), $\mathcal{U}$ the class of abstraction nodes (*i.e.*, unary nodes), and $\hat{\mathcal{P}}^{(i,k)}$ the class of unary-binary trees such that every leaf e can be labelled in $\min\{\ell(e) + i, k\}$ ways, where $\ell(v)$ denotes the de Bruijn level of v. The objects in $\hat{\mathcal{P}}^{(i,k)}$ may be seen as lambda-DAGs where the binding of each variable x may come from a binding unary node at most k De Bruijn levels above x, even if this means up to i De Bruijn levels above the root. Therefore, the class we are interested in is $\hat{\mathcal{P}}^{(0,k)}$

The classes $\hat{\mathcal{P}}^{(i,k)}$ can be specified by

$$\hat{\mathcal{P}}^{(k,k)} = k\mathcal{Z} + (\mathcal{A} \times \hat{\mathcal{P}}^{(k,k)} \times \hat{\mathcal{P}}^{(k,k)}) + (\mathcal{U} \times \hat{\mathcal{P}}^{(k,k)}),$$

and

$$\hat{\mathcal{P}}^{(i,k)} = i\mathcal{Z} + (\mathcal{A} \times \hat{\mathcal{P}}^{(i,k)} \times \hat{\mathcal{P}}^{(i,k)}) + (\mathcal{U} \times \hat{\mathcal{P}}^{(i+1,k)}) \qquad \text{for} \ \ i < k.$$

Translating into generating functions with z marking the size and u marking the number of leaves, we get

$$\hat{P}^{(k,k)}(z,u) = kzu + z\hat{P}^{(k,k)^2}(z,u) + z\hat{P}^{(k,k)}(z,u),$$

and

$$\hat{P}^{(i,k)}(z,u) = izu + z\hat{P}^{(i,k)^2}(z,u) + z\hat{P}^{(i+1,k)}(z,u),$$

which yields

$$\hat{P}^{(k,k)}(z,u) = \frac{1 - z - \sqrt{(1-z)^2 - 4z^2 ku}}{2z},$$

and

$$\hat{P}^{(i,k)}(z,u) = \frac{1 - \sqrt{1 - 4z(izu + z\hat{P}^{(i+1,k)}(z,u))}}{2z} = \frac{1 - \sqrt{1 - 4iz^2 u - 4z^2 \hat{P}^{(i+1,k)}(z,u)}}{2z},$$

for $i < k$.

This can be written in the form

$$\hat{P}^{(i,k)}(z,u) = \frac{1 - \mathbf{1}_{[i=k]} z - \sqrt{\hat{R}_{k-i+1,k}(z,u)}}{2z},$$

with

$$\hat{R}_{1,k}(z,u) = (1-z)^2 - 4kuz^2, \tag{1}$$

$$\hat{R}_{2,k}(z,u) = 1 - 4(k-1)z^2 u - 2z + 2z^2 + 2z\sqrt{\hat{R}_{1,k}(z,u)},$$

and

$$\hat{R}_{i,k}(z,u) = 1 - 4(k-i+1)z^2 u - 2z + 2z\sqrt{\hat{R}_{i-1,k}(z,u)}, \qquad \text{for} \ \ 3 \le i \le k+1. \tag{2}$$

Since the class $\hat{\mathcal{P}}^{(0,k)}$ is isomorphic to the class $\mathcal{G}_k$ of closed lambda-terms where all De Bruijn indices are not larger than k, we get for the corresponding bivariate generating function

$$G_k(z,u) = \hat{P}^{(0,k)}(z,u) = \frac{1 - \sqrt{\hat{R}_{k+1,k}(z,u)}}{2z}.$$

We will now perform a singularity analysis of these generating functions. Recall that we call a singularity dominant, if it is a singularity of smallest modulus. Notice further that the behaviour of a generating function near its dominant singularities determines the asymptotic behaviour of its coefficients.

From [8] we know that the dominant singularity of $G_k(z,1)$ comes from the innermost radicand only and consequently is of type $\frac{1}{2}$. Due to continuity arguments this implies that in a sufficiently small neighbourhood of $u = 1$ the dominant singularity $\hat{\rho}_k(u)$ of $G_k(z,u)$ comes also only from the innermost radicand, i.e., $\hat{R}_{1,k}(z,u)$, and is of type $\frac{1}{2}$. By calculating the smallest positive root of $\hat{R}_{1,k}(z,u)$ we get $\hat{\rho}_k(u) = \frac{1}{1+2\sqrt{ku}}$. Now we will determine the expansions of the radicands in a neighbourhood of the dominant singularity $\hat{\rho}_k(u)$.

Let us briefly sketch our approach: We know that the singularity $\hat{\rho}_k(u)$ is determined by the equation $\hat{R}_{1,k}(\hat{\rho}_k(u),u) = 0$, provided that u is sufficiently close to 1, and that it is likewise the dominant singularity of $G_k(z,u)$ (seen as function of z, as all functions in the context right now). Consequently, it is the dominant singularity of all $\hat{R}_{j,k}(z,u)$, for $j = 2,\ldots,k+1$. If we determine the local behaviour of $\hat{R}_{1,k}(z,u)$ near $z = \hat{\rho}_k(u)$, then we will be able to determine Puiseux expansions of all $\hat{R}_{j,k}(z,u)$ for $j = 2,\ldots,k+1$ at $z = \hat{\rho}_k(u)$. This will be done in Proposition 1. In particular, this gives us the Puiseux expansions of $G_k(z,u)$ from which we can derive the asymptotic behaviour of its coefficients by transfer theorems [20, 19]. This will be the task of Theorem 5 below.

It will then turn out that the shape of $\hat{\rho}_k(u)$ near $u = 1$ determines the characteristic function of the random variable "number of leaves", because $\hat{\rho}_k(u)$ depends on u in a nicely regular way. This characteristic function has then the shape of a so-called quasi-power involving the function $\hat{\rho}_k(u)$. Hwang's Quasi-Powers Theorem [26] will then do the job and yield a central limit theorem.

Proposition 1. *Let $\hat{\rho}_k(u)$ be the root of the innermost radicand $\hat{R}_{1,k}(z,u)$, i.e. $\hat{\rho}_k(u) = \frac{1}{1+2\sqrt{ku}}$, where u is in a sufficiently small neighbourhood of 1, i.e. $|u-1| < \delta$ for $\delta > 0$ sufficiently small. Then the equations*

$$\hat{R}_{1,k}(\hat{\rho}_k(u) - \epsilon, u) = \left(2 - 2\hat{\rho}_k(u) + 8ku\hat{\rho}_k(u)\right)\epsilon + \mathcal{O}(|\epsilon|^2), \tag{3}$$

and

$$\hat{R}_{j,k}(\hat{\rho}_k(u) - \epsilon, u) = c_j(u)\hat{\rho}_k(u)^2 + \frac{\sqrt{8\hat{\rho}_k(u)(1 - \hat{\rho}_k(u)^2)}}{\prod_{l=2}^{j}\sqrt{c_l(u)}} \cdot \sqrt{\epsilon} + \mathcal{O}(|\epsilon|), \tag{4}$$

for $2 \leq j \leq k+1$, with $c_1(u) = 1$ and $c_j(u) = 4(j-1)u - 1 + 2\sqrt{c_{j-1}(u)}$ for $2 \leq j \leq k+1$, hold for $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C} \setminus \mathbb{R}^-$, uniformly in u.

Proof. Using the Taylor expansion of $\hat{R}_{1,k}(z,u)$ around $\hat{\rho}_k(u)$ we obtain

$$\hat{R}_{1,k}(z,u) = \hat{R}_{1,k}(\hat{\rho}_k(u),u) + (z - \hat{\rho}_k(u))\frac{\partial}{\partial z}\hat{R}_{1,k}(\hat{\rho}_k(u),u) + \mathcal{O}((z - \hat{\rho}_k(u))^2).$$

Per definition the first summand $\hat{R}_{1,k}(\hat{\rho}_k(u),u)$ is equal to zero. Setting $z = \hat{\rho}_k(u) - \epsilon$ and using (1) we obtain the first claim of Proposition 1.

The next step is to compute an expansion of $\hat{R}_{j,k}(z,u)$ around $\hat{\rho}_k(u)$ for $2 \leq j \leq k+1$. Using the recursive relation (1) for $\hat{R}_{2,k}(z,u)$ and the formula $\hat{\rho}_k(u) = \frac{1}{1+2\sqrt{ku}}$ yields

$$\hat{R}_{2,k}(\hat{\rho}_k(u) - \epsilon, u) = (1 + 4u)\hat{\rho}_k(u)^2 + 2\hat{\rho}_k(u)\sqrt{2 - 2\hat{\rho}_k(u) + 8ku\hat{\rho}_k(u)}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

We set $c_2(u) := 1 + 4u$ and $d_2(u) := 2\hat{\rho}_k(u)\sqrt{2 - 2\hat{\rho}_k(u) + 8ku\hat{\rho}_k(u)}$ and assume that for $2 \leq j \leq k+1$ the equation $\hat{R}_{j,k}(\hat{\rho}_k(u) - \epsilon, u) = c_j(u)\hat{\rho}_k^2(u) + d_j(u)\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$ holds. Now we proceed by induction.

Observe that

$$\hat{R}_{j+1,k}(\hat{\rho}_k(u) - \epsilon, u) = 1 - 4(k-j)\hat{\rho}_k^2(u)u - 2\hat{\rho}_k(u) + 2\hat{\rho}_k(u)\sqrt{c_j(u)\hat{\rho}_k^2 + d_j(u)\sqrt{\epsilon}} + \mathcal{O}(|\epsilon|).$$

Expanding, using again $\hat{\rho}_k(u) = \frac{1}{1+2\sqrt{ku}}$, and simplifying yields

$$\hat{R}_{j+1,k}(\hat{\rho}_k(u) - \epsilon, u) = 4ju\hat{\rho}_k(u)^2 - \hat{\rho}_k(u)^2 + 2\hat{\rho}_k(u)^2\sqrt{c_j(u)} + \frac{d_j(u)}{\sqrt{c_j(u)}}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

Setting $c_{j+1}(u) := 4ju - 1 + 2\sqrt{c_j(u)}$ and $d_{j+1}(u) := \frac{d_j(u)}{\sqrt{c_j(u)}}$ for $2 \leq j \leq k$, we obtain
$\hat{R}_{j+1,k}(\hat{\rho}_k(u) - \epsilon, u) = c_{j+1}\hat{\rho}_k^2(u) + d_{j+1}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$.

Expanding $d_{j+1}(u)$, using its recursive relation and $d_2(u) = 2\hat{\rho}_k(u)\sqrt{2 - 2\hat{\rho}_k(u) + 8ku\hat{\rho}_k(u)}$, we get for $2 \leq j \leq k$

$$d_{j+1}(u) = \frac{2\hat{\rho}_k(u)\sqrt{2 - 2\hat{\rho}_k(u) + 8ku\hat{\rho}_k(u)}}{\prod_{l=2}^{j}\sqrt{c_l(u)}}.$$

Finally, we show that the $c_l(u)$'s are greater than zero in a neighbourhood of $u = 1$. By induction it can easily be seen that they are always positive for $u = 1$, since

$$c_1(1) = 1,$$

and assuming $c_{i-1}(1) < c_i(1)$ we get

$$c_{i+1}(1) = 4i - 1 + 2\sqrt{c_i(1)} > 4(i-1) + 4 - 1 + 2\sqrt{c_{i-1}(1)} = c_i(1) + 4.$$

Using continuity arguments we can see that the functions $c_l(u)$ have to be positive in a sufficiently small neighbourhood of $u = 1$ as well, which completes the proof of (4).   □

Theorem 5. *Let for any fixed k, $G_k(z, u)$ denote the bivariate generating function of the class of closed lambda-terms where all De Bruijn indices are at most k. Then the equation*

$$[z^n]G_k(z, u) = \sqrt{\frac{\sqrt{ku} + 2ku}{4\pi \prod_{l=2}^{k+1} c_l(u)}}(1 + 2\sqrt{ku})^n n^{-\frac{3}{2}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \quad for \ \ n \to \infty,$$

with $c_1(u) = 1$ and $c_j(u) = 4(j-1)u - 1 + 2\sqrt{c_{j-1}(u)}$, for $2 \leq j \leq k+1$, holds uniformly in u for $|u - 1| < \delta$, with $\delta > 0$ sufficiently small.

Proof. Using $G_k(z, u) = \frac{1 - \sqrt{\hat{R}_{k+1,k}(z,u)}}{2z}$ and (4), we get for $\epsilon \in \mathbb{C} \setminus \mathbb{R}^-$ with $|\epsilon| \longrightarrow 0$

$$G_k(\hat{\rho}_k(u) - \epsilon, u) = \frac{1 - \sqrt{c_{k+1}(u)}\hat{\rho}_k(u)}{2\hat{\rho}_k(u)} - \frac{d_{k+1}(u)}{4\hat{\rho}_k(u)^2\sqrt{c_{k+1}(u)}}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

Hence,

$$[z^n]G_k(z, u) = -\frac{d_{k+1}(u)\sqrt{\hat{\rho}_k(u)}}{4\hat{\rho}_k^2(u)\sqrt{c_{k+1}(u)}}[z^n]\sqrt{1 - \frac{z}{\hat{\rho}_k(u)}} + [z^n]\mathcal{O}(\hat{\rho}_k(u) - z).$$

Since $\hat{\rho}_k(u) = \frac{1}{1+2\sqrt{ku}}$ is of type $\frac{1}{2}$ and by plugging in the formula for $d_{k+1}(u) = \frac{2\hat{\rho}_k(u)\sqrt{2-2\hat{\rho}_k(u)+8ku\hat{\rho}_k(u)}}{\prod_{l=2}^{k}\sqrt{c_l(u)}}$, we obtain the desired result by applying the standard transfer theorems (see [20, 19]).   □

From [8, Theorem 1] we know the following result:

$$[z^n]G_k(z, 1) = \sqrt{\frac{\sqrt{k} + 2k}{4\pi \prod_{l=2}^{k+1} c_l(1)}}(1 + 2\sqrt{k})^n n^{-\frac{3}{2}}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \quad \text{as } n \to \infty, \qquad (5)$$

with $c_l(u)$ defined as in Proposition 1.

Now we want to apply the well-known Quasi-Powers Theorem.

Theorem 6 (Quasi-Powers Theorem, [26])**.** *Let X_n be a sequence of random variables with the property that*

$$\mathbb{E}u^{X_n} = A(u)B(u)^{\lambda_n}\left(1 + \mathcal{O}\left(\frac{1}{\phi_n}\right)\right)$$

holds uniformly in a complex neighbourhood of $u = 1$, where $\lambda_n \to \infty$ and $\phi_n \to \infty$, and $A(u)$ and $B(u)$ are analytic functions in a neighbourhood of $u = 1$ with $A(1) = B(1) = 1$. Set $\mu = B'(1)$ and $\sigma^2 = B''(1) + B'(1) - B'(1)^2$. If $\sigma^2 \neq 0$, then

$$\frac{X_n - \mathbb{E}X_n}{\sqrt{\mathbb{V}X_n}} \to \mathcal{N}(0,1),$$

with $\mathbb{E}X_n = \mu\lambda_n + A'(1) + \mathcal{O}(1/\phi_n))$ and $\mathbb{V}X_n = \sigma^2\lambda_n + A''(1) + A'(1) - A'(1)^2 + \mathcal{O}(1/\phi_n))$.

Using Theorem 5 and (5), we get for $n \longrightarrow \infty$

$$\mathbb{E}u^{X_n} = \frac{[z^n]G_k(z,u)}{[z^n]G_k(z,1)} = \left(\frac{1 + 2\sqrt{ku}}{1 + 2\sqrt{k}}\right)^n \sqrt{\frac{\sqrt{ku} + 2ku}{2k + \sqrt{k}} \prod_{j=2}^{k+1} \frac{c_j(1)}{c_j(u)}} \left(1 + O\left(\frac{1}{n}\right)\right),$$

where $c_1(u) = 1$ and $c_j(u) = 4ju - 4u - 1 + 2\sqrt{c_{j-1}(u)}$.

Thus, all assumptions for the Quasi-Powers Theorem are fulfilled, and we get that the number of leaves in closed lambda-terms with De Bruijn indices at most k is asymptotically normally distributed with

$$\mathbb{E}X_n \sim \frac{k}{\sqrt{k} + 2k}n, \quad \text{and} \quad \mathbb{V}X_n \sim \frac{k^2}{2\sqrt{k}(\sqrt{k} + 2k)^2}n, \quad \text{as } n \to \infty,$$

and therefore Theorem 1 is shown.

4. Total number of leaves in random lambda-terms with bounded number of De Bruijn levels

This section is devoted to the enumeration of leaves in closed lambda-terms with a bounded number of De Bruijn levels. As in [8] let us denote by $\mathcal{P}^{(i,k)}$ the class of unary-binary trees that contain at most $k - i$ De Bruijn levels and each leaf e can be coloured with one out of $i + l(e)$ colors, where $l(e)$ denotes the De Bruijn level in which the respective leaf is located. These classes can be specified by

$$\mathcal{P}^{(k,k)} = k\mathcal{Z} + (\mathcal{A} \times \mathcal{P}^{(k,k)} \times \mathcal{P}^{(k,k)}),$$

and

$$\mathcal{P}^{(i,k)} = i\mathcal{Z} + (\mathcal{A} \times \mathcal{P}^{(i,k)} \times \mathcal{P}^{(i,k)}) + (\mathcal{U} \times \mathcal{P}^{(i+1,k)}) \quad \text{for } i < k.$$

By translating into generating functions we get

$$P^{(k,k)}(z,u) = kzu + zP^{(k,k)^2}(z,u),$$

and

$$P^{(i,k)}(z,u) = izu + zP^{(i,k)^2}(z,u) + zP^{(i+1,k)}(z,u) \quad \text{for } i < k.$$

Solving yields

$$P^{(k,k)}(z,u) = \frac{1 - \sqrt{1 - 4kz^2u}}{2z},$$

and

$$P^{(i,k)}(z,u) = \frac{1 - \sqrt{1 - 4iz^2u - 4z^2P^{(i+1,k)}}}{2z} \quad \text{for } i < k.$$

This can be written as

$$P^{(i,k)}(z,u) = \frac{1 - \sqrt{R_{k-i+1,k}(z,u)}}{2z},$$

where

$$R_{1,k}(z,u) = 1 - 4kz^2u, \tag{6}$$

and

$$R_{i,k}(z,u) = 1 - 4(k-i+1)z^2 u - 2z + 2z\sqrt{R_{i-1,k}(z,u)}, \quad \text{for } 2 \le i \le k+1. \tag{7}$$

For the bivariate generating function of closed lambda-terms with at most k De Bruijn levels we get

$$H_k(z,u) = P^{(0,k)}(z,u) = \frac{1 - \sqrt{R_{k+1,k}(z,u)}}{2z}.$$

Thus, the generating function consists again of $k+1$ nested radicals, but as stated in Section 2, the counting sequence of this class of lambda-terms shows a very unusual behaviour. The type of the dominant singularity of the generating function changes when the imposed bound equals N_j. Thus, the subexponential term in the asymptotics of the counting sequence changes. The following result has been shown in [8]:

Theorem 7 ([8, Theorem 3]). *Let $(u_i)_{i \ge 0}$ and $(N_i)_{i \ge 0}$ be the integer sequences defined in Definition 4 and let $H_k(z,1)$ be the generating function of the class of closed lambda-terms with at most k De Bruijn levels. Then the following asymptotic relations hold*

(i) *If there exists $j \ge 0$ such that $N_j < k < N_{j+1}$, then there exists a constant h_k such that*

$$[z^n]H_k(z,1) \sim h_k n^{-3/2}\rho_k(1)^{-n}, \ as \ n \to \infty.$$

(ii) *If there exists j such that $k = N_j$, then*

$$[z^n]H_k(z,1) \sim h_k n^{-5/4}\rho_k(1)^{-n} = h_k n^{-5/4}(2u_j)^n, \ as \ n \to \infty.$$

Thus, in order to investigate structural properties of this class of lambda-terms we perform a distinction of cases whether the bound k is an element of the sequence $(N_i)_{i \ge 0}$ or not.

4.1. **The case $N_j < k < N_{j+1}$.** From [8] we know that in this case the dominant singularity of the generating function $H_k(z,1)$ comes from the $(j+1)$-th radicand $R_{j+1,k}$ and is of type $\frac{1}{2}$. As in the previous section we can again use continuity arguments to guarantee that sufficiently close to $u = 1$ the dominant singularity $\rho_k(u)$ of $H_k(z,u)$ comes from the $(j+1)$-th radicand $R_{j+1,k}(z,u)$ and is of type $\frac{1}{2}$. Now we will determine the expansions of the radicands in a neighbourhood of the dominant singularity.

Proposition 2. *Let $\rho_k(u)$ be the dominant singularity of $H_k(z,u)$, where u is in a sufficiently small neighbourhood of 1, i.e. $|u - 1| < \delta$ for $\delta > 0$ sufficiently small. Then the expansions*

(i) $\forall i < j+1$ *(inner radicands)* $: R_{i,k}(\rho_k(u) - \epsilon, u) = R_{i,k}(\rho_k(u), u) + \mathcal{O}(|\epsilon|)$

(ii) $R_{j+1,k}(\rho_k(u) - \epsilon, u) = \gamma_{j+1}(u)\epsilon + \mathcal{O}(|\epsilon|^2)$, *with* $\gamma_{j+1}(u) = -\frac{\partial}{\partial z}R_{j+1,k}(\rho_k(u), u)$

(iii) $\forall i > j+1$ *(outer radicands)* $: R_{i,k}(\rho_k(u) - \epsilon, u) = a_i(u) + b_i(u)\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$, *with* $a_{i+1}(u) = 1 - 4(k-i)\rho_k(u)^2 u - 2\rho_k(u) + 2\rho_k(u)\sqrt{a_i(u)}$, *and* $b_{i+1}(u) = \frac{b_i(u)\rho_k(u)}{\sqrt{a_i(u)}}$ *for* $j+2 \le i \le k$,

with $a_{j+2}(u) = 1 - 4(k-j-1)\rho_k(u)^2 u - 2\rho_k(u)$ *and* $b_{j+2}(u) = 2\rho_k(u)\sqrt{\gamma_{j+1}(u)}$,

hold for $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C} \setminus \mathbb{R}^-$, uniformly in u.

Proof. (i) The first equation (for $i < j+1$) follows immediately by Taylor expansion around $\rho_k(u)$ and setting $z = \rho_k(u) - \epsilon$.

(ii) The equation for $i = j+1$ follows analogously to the first case, knowing that $R_{j+1,k}(z,u)$ cancels for $z = \rho_k(u)$.

(iii) The next step is to expand $R_{i,k}(z,u)$ around $\rho_k(u)$ for $i > j+1$. From the second claim of Proposition 2 and from the recurrence relation (7) for $R_{i,k}(z,u)$ it results

$$R_{j+2,k}(\rho_k(u) - \epsilon, u) = 1 - 4(k-j-1)\rho_k(u)^2 u - 2\rho_k(u) + 2\rho_k(u)\sqrt{\gamma_{j+1}(u)}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

We set $a_{j+2}(u) := 1 - 4(k-j-1)\rho_k^2(u)u - 2\rho_k(u)$ and $b_{j+2}(u) := 2\rho_k(u)\sqrt{\gamma_{j+1}(u)}$. Now we proceed by induction. Assume $R_{i,k}(\rho_k(u) - \epsilon, u) = a_i(u) + b_i(u)\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$. We have just checked that it holds for $i = j+2$. Now we perform the induction step $i \mapsto i+1$.

Using the recursion (7) for $R_{i,k}$ and plugging in the expansion $a_i(u) + b_i(u)\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$ for $R_{i,k}(\rho_k(u) - \epsilon, u)$ yields

$$R_{i+1,k}(\rho_k(u) - \epsilon, u) = 1 - 4(k-i)\rho_k(u)^2 u - 2\rho_k(u) + 2\rho_k(u)\sqrt{a_i(u)} + \frac{b_i(u)\rho_k(u)}{\sqrt{a_i(u)}}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

Setting $\forall i \geq j + 2$ $a_{i+1}(u) := 1 - 4(k-i)\rho_k^2(u)u - 2\rho_k(u) + 2\rho_k(u)\sqrt{a_i(u)}$ and $b_{i+1}(u) := \frac{b_i(u)\rho_k(u)}{\sqrt{a_i(u)}}$, we obtain

$$R_{i+1,k}(\rho_k(u) - \epsilon, u) = a_{i+1}(u) + b_{i+1}(u)\sqrt{\epsilon} + \mathcal{O}(|\epsilon|).$$

Expanding $b_i(u)$, using its recursive relation and $b_{j+2}(u) = 2\rho_k(u)\sqrt{\gamma_{j+1}(u)}$ we get for $i > j + 1$

$$b_i(u) = \frac{2\rho_k^{i-j}(u)\sqrt{\gamma_{j+1}(u)}}{\prod_{l=j+1}^{i-1}\sqrt{a_l(u)}}.$$

$\square$

We know that for sufficiently large i the sequence u_i, defined in Definition 4, is given by $u_i = \lfloor \chi^{2^i} \rfloor$, with $\chi \approx 1.36660956\ldots$ (see [8, Lemma 18]). Therefore we have $N_j \sim u_j^2 \sim \chi^{2^j 2}$ and $N_j < k < N_{j+1} = O(N_j^2)$, which gives $j = \Theta(\log\log k)$. This implies that $j + 1 < k + 1$, i.e., that the dominant singularity $\rho_k(u)$ cannot come from the outermost radical.

Remark. Obviously the same is true for the case $k = N_j$. Thus, the dominant singularity never comes from the outermost radical.

Using Proposition 2 and $H_k(z, u) = \frac{1}{2z}(1 - \sqrt{R_{k+1,k}(z, u)})$ we get

$$H_k(\rho_k(u) - \epsilon, u) = \frac{1 - \sqrt{a_{k+1}(u)}}{2\rho_k(u)} - \frac{b_{k+1,k}(u)}{4\rho_k(u)\sqrt{a_{k+1}(u)}}\sqrt{\epsilon} + \mathcal{O}(|\epsilon|),$$

which yields

$$[z^n]H_k(z, u) = h_k(u)\rho_k(u)^{-n}\frac{n^{-\frac{3}{2}}}{\Gamma(-\frac{1}{2})}\left(1 + O\left(\frac{1}{\sqrt{n}}\right)\right), \tag{8}$$

with

$$h_k(u) = -\frac{b_{k+1}(u)\sqrt{\rho_k(u)}}{4\rho_k(u)\sqrt{a_{k+1}(u)}}.$$

Taking a look at the recursive definitions of $a_i(u)$ and $b_i(u)$ (see Proposition 2), it can easily be seen that these functions are not equal to zero in a sufficiently small neighbourhood of $u = 1$. We know that $a_{j+2}(1)$ is positive, since

$$a_{j+2}(1) = 1 - 4(k-j-1)\rho_k(1)^2 - 2\rho_k(1) = 1 - 4(k-j)\rho_k(1)^2 - 2\rho_k(1) + 4\rho_k^2,$$

and $1 - 4(k-j)\rho_k(1)^2 - 2\rho_k(1) > 0$ (see [8]). By induction we can show that the sequence $a_i := a_i(1)$ is monotonically increasing. Let us assume that $a_{i-1} < a_i$, then we get

$$a_{i+1} > 1 - 4(k-i)\rho_k(1)^2 - 2\rho_k(1) + 2\rho_k(1)\sqrt{a_i}$$
$$> 1 - 4(k-i+1)\rho_k(1)^2 - 2\rho_k(1) + 2\rho_k(1)\sqrt{a_{i-1}} + 4\rho_k(1)^2 > a_i + 4\rho_k(1)^2.$$

It is obvious that if $b_{j+2} := b_{j+2}(1)$ is non-zero, than all the b_i's, which are defined by

$$b_i = \frac{\rho_k(1)b_{i-1}}{a_{i-1}},$$

are non-zero. In order to prove that $b_{j+2} = 2\rho_k(1)\sqrt{-\frac{\partial}{\partial z}R_{j+1,k}(\rho_k(1), 1)}$ is non-zero, we also proceed by induction. Since

$$R_{1,k}(z, 1) = 1 - 4kz^2,$$

we can see that $\frac{\partial}{\partial z} R_{1,k}(\rho_k(1), 1) < 0$, and assuming $\frac{\partial}{\partial z} R_{i,k}(\rho_k(1), 1) < 0$ and using

$$\frac{\partial}{\partial z} R_{i+1,k}(z, 1) = -8(k-i)z - 2 + 2\sqrt{R_{i,k}(z, 1)} + \frac{z}{\sqrt{R_{i,k}(z, 1)}} \frac{\partial}{\partial z} R_i(z, 1),$$

we proved that all b_i's are non-zero. Thus, we get that $h_k(u) \neq 0$.

Using (8) and Theorem 7 we get for $n \longrightarrow \infty$

$$\frac{[z^n]H_k(z, u)}{[z^n]H_k(z, 1)} = \frac{h_k(u)}{h_k\Gamma(-1/2)} \left(\frac{\rho_k(1)}{\rho_k(u)}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right). \tag{9}$$

Assuming that $\sigma^2 = B''(1) + B'(1) - B'(1)^2 \neq 0$ with $B(u) = \frac{\rho_k(1)}{\rho_k(u)}$ we can apply the Quasi-Powers Theorem. As stated in Section 2 the proof of this assumption appears to be quite difficult, since there is only very little known about the function $\rho_k(u)$. However, it seems very likely that this condition will be fulfilled for arbitrary $k \in (N_j, N_{j+1})$, so that the Quasi-Powers Theorem can be applied and we get that the number of leaves in lambda-terms with bounded number of De Bruijn levels is asymptotically normally distributed with asymptotic mean μn and variance $\sigma^2 n$, respectively, where $\mu = B'(1)$ and $\sigma^2 = B''(1) + B'(1) - B'(1)^2$, with $B(u) = \frac{\rho_k(1)}{\rho_k(u)}$.

bound k	$j+1$	$B''(1) + B'(1) - B'(1)^2$	$B'(1)$
1	**2**	**0**	**0**
2	2	0.0385234386	0.4381229337
3	2	0.0210625856	0.4414407371
4	2	0.0167136805	0.4463973717
5	2	0.0148700270	0.4504258849
6	2	0.0138224393	0.4536185043
7	2	0.0131157948	0.4561987871
8	**3**	**0.0125868052**	**0.4583333333**
9	3	0.0582322465	0.4566104777
10	3	0.0470481360	0.4560418340
11	3	0.0396601986	0.4560810348
12	3	0.0345090124	0.4564489368
$\vdots$	$\vdots$	$\vdots$	$\vdots$
133	3	0.0077469541	0.4821900098
134	3	0.0077234960	0.4822482745
135	**4**	**0.0077002803**	**0.4823059361**
136	4	0.0132855719	0.4823515285
137	4	0.0131816901	0.4823968564
138	4	0.0130800422	0.4824419195
139	4	0.0129805564	0.4824867175

TABLE 1. Table summarizing the coefficients occurring in the variance and the mean for some initial values of k.

4.2. **The case $k = N_j$.** We know from [8] that in the case $k = N_j$ both radicands $R_{j,k}(z, 1)$ and $R_{j+1,k}(z, 1)$ vanish simultaneously and the dominant singularity is therefore of type $\frac{1}{4}$. This is not true for the radicands $R_{j,k}(z, u)$ and $R_{j+1,k}(z, u)$ when u is in a neighbourhood of 1. Thus, we have a discontinuity at $\rho_k(1)$, which is why we do not get any uniform expansions of the radicands in a neighbourhood of $\rho_k(1)$.

In order to overcome this problem we proceed as follows (see Figure 5 for a sketch of the idea of the proof): First, we show that the dominant singularity of the generating function $H_k(z, 1+\epsilon)$ comes solely from the radicand $R_{j,k}(z, 1+\epsilon)$ (*cf.* Lemma 1). Then we investigate the expansions of the radicands thoroughly for $u = 1 + \frac{s}{\sqrt{n}}$ in a neighbourhood of $z = \rho_k(1)$ with radius $\frac{t}{n}$, where s and t are both bounded complex numbers (*cf.* Lemma 2). This approach of choosing the considered neighbourhoods of $z = \rho_k(1)$ and $u = 1$ to be dependent on each other constitutes the main idea of

the applied method. By use of Cauchy's coefficient formula we are then able to obtain an asymptotic expression for the n-th coefficient of the generating function $H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right)$ by choosing a suitable integration contour (*cf.* Proposition 3). Finally, we show that the characteristic function of the random variable counting the total number of variables in a random lambda-term with at most k De Bruijn levels tends to the characteristic function of the normal distribution as the size tends to infinity (*cf.* Lemma 3). For convenience, we will subsequently use the abbreviation $\rho_k := \rho_k(1)$.
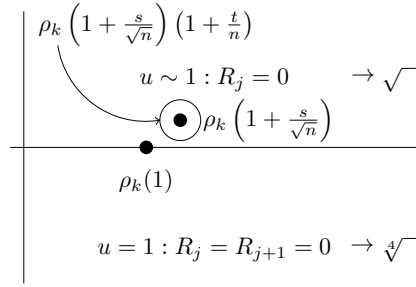

FIGURE 5. Sketch of the idea of the proof.

Lemma 1. *For $u = 1 + \epsilon$ with $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C} \backslash \mathbb{R}_0^-$, the dominant singularity $\rho_k(u) = \rho_k(1+\epsilon)$ of the bivariate generating function $H_k(z, 1 + \epsilon)$ comes from the j-th radicand $R_{j,k}(z, u)$.*

Proof. Setting $u = 1 + \epsilon$, expanding $\rho_k(u)$ around 1 and plugging into the recursive definition of the radicands yields

$$R_{j,k}\left(\rho_k(1+\epsilon), 1+\epsilon\right) = 1 - 4(k-j+1)(\rho_k^2 + 2\rho_k\rho_k'\epsilon + (\rho_k'^2 + 2\rho_k\rho_k'')\epsilon^2 + \rho_k^2\epsilon)$$
$$- (2\rho_k + 2\rho_k'\epsilon + 2\rho_k''\epsilon^2)\left(1 - \sqrt{R_{j-1,k}\left(\rho_k(1+\epsilon), 1+\epsilon\right)}\right) + \mathcal{O}(|\epsilon|).$$

Using $1 - 4(k-j)\rho_k^2 - 2\rho_k = 0$ and $\sqrt{R_{j-1,k}\left(\rho_k(1+\epsilon), 1+\epsilon\right)} = \sqrt{R_{j-1,k}\left(\rho_k, 1\right) + \mathcal{O}(|\epsilon|)} = 2\rho_k + \mathcal{O}(|\epsilon|)$, which are both shown in [8], we get

$$R_{j,k}\left(\rho_k(1+\epsilon), 1+\epsilon\right) = -4(k-j+1)(2\rho_k\rho_k'\epsilon + (\rho_k'^2 + 2\rho_k\rho_k'')\epsilon^2)$$
$$- (2\rho_k'\epsilon + 2\rho_k''\epsilon^2)(1 - 2\rho_k + \mathcal{O}(|\epsilon|)) + \mathcal{O}(|\epsilon|).$$

Thus, $R_{j,k}\left(\rho_k(1+\epsilon), 1+\epsilon\right) = \Theta(|\epsilon|)$.

Using this result and again the recursive definition of the radicands results in

$$R_{j+1,k}\left(\rho_k(u), 1+\epsilon\right) = 2\sqrt{R_{j,k}(\rho_k(u), 1+\epsilon)} + \mathcal{O}(|\epsilon|) = \Theta(\sqrt{|\epsilon|}).$$

Thus, we see that $|R_{j+1,k}(\rho_k(u), u)| \gg |R_{j,k}(\rho_k(u), u)|$ in a neighbourhood of $u = 1$, which implies that the dominant singularity has to come from the j-th radicand, *i.e.* $R_{j,k}(\rho_k(u), u) = 0$ for u being sufficiently close to 1. $\square$

Lemma 2. *Let $z = \rho_k(u) = \rho_k(1 + \frac{s}{\sqrt{n}})$ be the dominant singularity of the bivariate generating function $H_k(z, 1 + \frac{s}{\sqrt{n}})$ with bounded $s \in \mathbb{C}$. Then, as $n \longrightarrow \infty$,*

(i) $R_{j,k}\left(\rho_k(u)\left(1 + \frac{t}{n}\right), 1 + \frac{s}{\sqrt{n}}\right) = \frac{1}{n}p_j(t) + \mathcal{O}\left(\frac{1}{n^{3/2}}\right),$
   with $p_j(t) := -8t(k-j+1)\rho_k^2 - 2\rho_k t + 4\rho_k^2 t + 2t\rho_k f(\frac{t}{n})$ where f is an analytic function around 0;

(ii) $R_{j+1,k}\left(\rho_k(u)\left(1 + \frac{t}{n}\right), 1 + \frac{s}{\sqrt{n}}\right) = \frac{1}{\sqrt{n}}p_{j+1}(s,t) + \mathcal{O}\left(\frac{1}{n}\right),$
   where $p_{j+1}(s,t) = 2\rho_k\sqrt{p_j(t)} - 4(k-j)(2\rho_k\rho_k's + \rho_k^2 s) - 2\rho_k's;$

(iii) $R_{i,k}\left(\rho_k(u)\left(1+\frac{t}{n}\right),1+\frac{s}{\sqrt{n}}\right)=\hat{C}_i+\frac{1}{\sqrt[4]{n}}p_i(s,t)+O\left(\frac{1}{\sqrt{n}}\right)$ for $i\geq j+2$,

where $\hat{C}_i$ are constants and p_i are analytic functions in the variables s and t.

Proof. We start with setting $u=1+\frac{s}{\sqrt{n}}$ and $z=\rho_k(u)(1+\frac{t}{n})$ with bounded $s,t\in\mathbb{C}$ (*cf.* Figure 5), which results in

$$R_{j+1,k}\left(\rho_k(u)\left(1+\frac{t}{n}\right),1+\frac{s}{\sqrt{n}}\right)=$$
$$1-4(k-j)\rho_k(u)^2\left(1+\frac{t}{n}\right)^2\left(1+\frac{s}{\sqrt{n}}\right)-2\rho_k(u)\left(1+\frac{t}{n}\right)\left(1-\sqrt{R_{j,k}}\right),$$

and

$$R_{j,k}\left(\rho_k(u)\left(1+\frac{t}{n}\right),1+\frac{s}{\sqrt{n}}\right)=$$
$$1-4(k-j+1)\rho_k(u)^2\left(1+\frac{t}{n}\right)^2\left(1+\frac{s}{\sqrt{n}}\right)-2\rho_k(u)\left(1+\frac{t}{n}\right)\left(1-\sqrt{R_{j-1,k}}\right),$$

where the radicand in the square root in the last bracket of both equations is of course also evaluated at $(z,u)=\left(\rho_k(1+\frac{s}{\sqrt{n}})(1+\frac{t}{n}),1+\frac{s}{\sqrt{n}}\right)$, but we will omit this notation from now on to ensure a simpler reading, *i.e.*, subsequently we will write $R_{i,k}$ instead of $R_{i,k}\left(\rho_k(1+\frac{s}{\sqrt{n}})\left(1+\frac{t}{n}\right),1+\frac{s}{\sqrt{n}}\right)$.

Expanding $\rho_k(1+\frac{s}{\sqrt{n}})$ around 1 and using the recursive definition for the radicands yields

$$R_{j,k}=1-4(k-j+1)\left(\rho_k^2+2\rho_k\rho_k'\frac{s}{\sqrt{n}}+(\rho_k'^2+2\rho_k\rho_k'')\frac{s^2}{n}+\rho_k^2\frac{s}{\sqrt{n}}+2\rho_k\rho_k'\frac{s^2}{n}+\rho_k^2\frac{2t}{n}\right)$$
$$-2\left(\rho_k+\rho_k'\frac{s}{\sqrt{n}}+\rho_k''\frac{s^2}{2n}+\rho_k\frac{t}{n}\right)\left(1-\sqrt{R_{j-1,k}}\right)+\mathcal{O}\left(\frac{1}{n^{3/2}}\right). \tag{10}$$

From Lemma 1 we know that for u in a sufficiently small vicinity of 1 the dominant singularity of $H_k(z,u)$ comes from the j-th radicand, *i.e.* $R_{j,k}\left(\rho_k(u),u\right)=0$. Expanding $R_{j,k}\left(\rho_k(1+\frac{s}{\sqrt{n}}),1+\frac{s}{\sqrt{n}}\right)$ this yields

$$1-4(k-j+1)\left(\rho_k^2+2\rho_k\rho_k'\frac{s}{\sqrt{n}}+(\rho_k'^2+2\rho_k\rho_k'')\frac{s^2}{n}+\rho_k^2\frac{s}{\sqrt{n}}+2\rho_k\rho_k'\frac{s^2}{n}\right)$$
$$-2\left(\rho_k+\rho_k'\frac{s}{\sqrt{n}}+\rho_k''\frac{s^2}{2n}\right)\left(1-\sqrt{R_{j-1,k}\left(\rho_k\left(1+\frac{s}{\sqrt{n}}\right),1+\frac{s}{\sqrt{n}}\right)}\right)+\mathcal{O}\left(\frac{1}{n^{3/2}}\right)=0.$$

Thus, Equation (10) simplifies to

$$R_{j,k}=-4(k-j+1)\rho_k^2\frac{2t}{n}-2\rho_k\frac{t}{n}+4\rho_k^2\frac{t}{n}+2\rho_k\frac{t}{n}f\left(\frac{t}{n}\right)+\mathcal{O}\left(\frac{1}{n^{3/2}}\right), \tag{11}$$

where $\frac{t}{n}f\left(\frac{t}{n}\right)=\sqrt{R_{j-1,k}}-\sqrt{R_{j-1,k}\left(\rho_k(1+\frac{s}{\sqrt{n}}),1+\frac{s}{\sqrt{n}}\right)}$. Notice that $f(x)$ is analytic around $x=0$.

Therefore, the proof of (*i*) is finished.

Proceeding equivalently for $R_{j+1,k}$ results in

$$R_{j+1,k}=\frac{1}{\sqrt{n}}\left(-4(k-j)(2\rho_k\rho_k's+\rho_k^2s)-2\rho_k's\right)+2\rho_k\sqrt{R_{j,k}}+O\left(\frac{1}{n}\right).$$

Finally, to complete the proof of the second statement of the assertion we simply have to replace $R_{j,k}$ by the right-hand side of (11). Going one step further leads to

$$R_{j+2,k}=\hat{C}_{j+2}+\frac{1}{\sqrt[4]{n}}p_{j+2}(s,t)+O\left(\frac{1}{\sqrt{n}}\right),$$

with $\hat{C}_{j+2} := 4\rho_k^2$ and $p_{j+2}(s,t) := 2\rho_k\sqrt{p_{j+1}(s,t)}$, where $p_{j+1}(s,t)$ is defined as in Lemma 2.

Now we proceed by induction. Therefore we assume that $R_{i,k} = \hat{C}_i + \frac{1}{\sqrt[4]{n}}p_i(s,t) + O\left(\frac{1}{\sqrt{n}}\right)$ with $i \geq j + 2$. Thus, we get

$$R_{i+1,k} = 1 - 4(k-i)\left(\rho_k^2 + 2\rho_k\rho_k'\frac{s}{\sqrt{n}} + (\rho_k'^2 + 2\rho_k\rho_k'')\frac{s^2}{n} + \rho_k^2\frac{s}{\sqrt{n}} + 2\rho_k\rho_k'\frac{s^2}{n} + \rho_k^2\frac{2t}{n}\right)$$
$$- 2\left(\rho_k + \rho_k'\frac{s}{\sqrt{n}} + \rho_k''\frac{s^2}{2n} + \rho_k\frac{t}{n}\right)\left(1 - \sqrt{R_{i,k}}\right) + \mathcal{O}\left(\frac{1}{n^{3/2}}\right).$$

Inserting the induction hypothesis and simplifying yields

$$R_{i+1,k} = 4(i-j)\rho_k^2 + 2\rho_k\sqrt{\hat{C}_i} + \frac{1}{\sqrt[4]{n}}\frac{\rho_k p_i(s,t)}{\sqrt{\hat{C}_i}} + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Setting $\hat{C}_{i+1} := 4(i-j)\rho_k^2 + 2\rho_k\sqrt{\hat{C}_i}$ and $p_{i+1}(s,t) := \frac{\rho_k}{\sqrt{\hat{C}_i}}p_i(s,t)$ completes the proof. $\qquad\square$

Proposition 3. *Let $H_k(z,u)$ be the bivariate generating function of the class of closed lambda-terms with at most k De Bruijn levels. Then the n-th coefficient of $H_k(z, 1 + \frac{s}{\sqrt{n}})$ with bounded $s \in \mathbb{C}$ is given by*

$$[z^n]H_k(z, 1 + \frac{s}{\sqrt{n}}) = C_k(s)\rho_k^{-n}n^{-\frac{5}{4}}\left(1 + O\left(n^{-\frac{3}{4}}\right)\right), \quad as\ n \longrightarrow \infty,$$

with a constant $C_k(s) \neq 0$.

Proof. Let us remember that $H_k(z, 1 + \frac{s}{\sqrt{n}}) = \frac{1 - \sqrt{R_{k+1,k}(z, 1 + \frac{s}{\sqrt{n}})}}{2z}$. Thus, with the well-known Cauchy coefficient formula we get

$$[z^n]H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right) = \frac{1}{2i\pi}\int_\gamma \frac{H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right)}{z^{n+1}}\,\mathrm{d}z$$

$$= \frac{1}{2i\pi}\int_\gamma \frac{1 - \sqrt{R_{k+1,k}\left(z, 1 + \frac{s}{\sqrt{n}}\right)}}{2z^{n+2}}\,\mathrm{d}z,$$

where γ encircles the dominant singularity $\rho_k(u)$ as depicted in Figure 6. We denote the small Hankel-like part of the integration contour γ that contributes the main part of the asymptotics by γ_H (*cf.* Figure 6). The curve γ_H encircles $\rho_k(u)$ at a distance $\frac{1}{n}$ and its straight parts (that lead into the direction $\rho_k(u) \cdot \infty$) have the length $\frac{\log^2(n)}{n}$. On $\gamma \setminus \gamma_H$ we have $|z| = |\rho_k(u)|\left|1 + \frac{\log^2(n)}{n} + \frac{i}{n}\right|$. This enables use to estimate the contribution of $\gamma \setminus \gamma_H$, which turns out to be exponentially small. Next, we use the transformation $z = \rho(u)\left(1 + \frac{t}{n}\right)$, which changes the integration contour γ_H to $\tilde{\gamma}_H$. On the new contour $\tilde{\gamma}_H$ the integrand is now represented in a way that Lemma 2 becomes directly applicable. Summarizing all those arguments, we know now that there exisits a $K > 0$ such that

$$[z^n]H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right) = \frac{1}{2i\pi}\int_{\tilde{\gamma}_H} \frac{1 - \sqrt{\hat{C}_{k+1} + \frac{1}{\sqrt[4]{n}}p_{k+1}(s,t) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)}}{2\rho^{n+1}e^t n}\,\mathrm{d}t + \mathcal{O}\left(e^{-K\log^2(n)}\right)$$

(12)

$$= \frac{1}{2i\pi}\int_{\tilde{\gamma}_H} \frac{1 - \sqrt{\hat{C}_{k+1}} - \frac{1}{2\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}p_{k+1}(s,t) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)}{2\rho_k^{n+1}e^t n}\,\mathrm{d}t + \mathcal{O}\left(e^{-K\log^2(n)}\right).$$

(13)

Now, let us observe how the function $p_{k+1}(s,t)$ looks like by using the recursive definition $p_{i+1}(s,t) = \frac{\rho_k}{\sqrt{\hat{C}_i}}p_i(s,t)$ and $p_{j+2}(s,t) = 2\rho_k\sqrt{2\rho_k\sqrt{p_j(t)} + q(s)}$, with a polynomial $q(s) =$
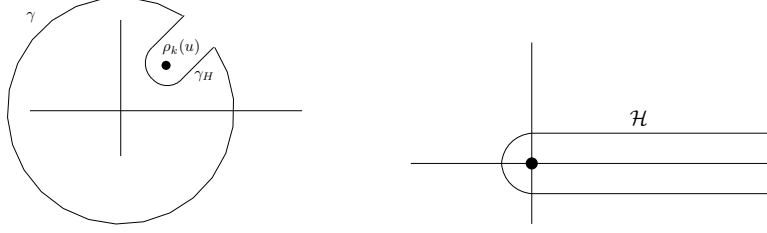
FIGURE 6. The integral contours γ and $\mathcal{H}$.

$s\left(-4(k-j)(2\rho_k\rho'_k + \rho_k^2) - 2\rho'_k\right)$ that is linear in s. Thus, $p_{k+1}(s,t) = D \cdot p_{j+2}(s,t)$ with a constant D. Inserting this into (13) and splitting the integral yields

$$[z^n]H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right) =$$

$$\frac{\rho_k^{-n}}{4i\pi\rho_k n}\left(\int_{\tilde{\gamma}_H}\left(1 - \sqrt{\hat{C}_{k+1}}\right)e^{-t}\,dt - \int_{\tilde{\gamma}_H}\frac{De^{-t}}{2\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}\sqrt{2\rho_k\sqrt{p_j(t)} + q(s)}\,dt + \int_{\tilde{\gamma}_H}\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)e^{-t}\,dt\right).$$

The first integral is zero and the third integral contributes $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$. Thus, the main part of the asymptotics results from the second integral: There are some constants $A(s)$ and $B(s)$ such that

$$-\int_{\tilde{\gamma}_H}\frac{De^{-t}}{\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}\sqrt{2\rho_k\sqrt{p_j(t)} + q(s)}\,dt = -\int_{\tilde{\gamma}_H}\frac{De^{-t}}{\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}\sqrt{A(s)t + B(s) + \mathcal{O}\left(\frac{\log^4(n)}{n}\right)}\,dt$$

$$= -\int_{\tilde{\gamma}_H}\frac{De^{-t}}{\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}\sqrt{A(s)t + B(s)}\,dt + \mathcal{O}\left(\frac{\log^6(n)}{n}\right)$$

$$= -\int_{\mathcal{H}}\frac{De^{-t}}{\sqrt[4]{n}\sqrt{\hat{C}_{k+1}}}\sqrt{A(s)t + B(s)}\,dt + \mathcal{O}\left(e^{-\tilde{K}\log^2(n)}\right)$$

$$\sim \tilde{C}(s)\frac{1}{\sqrt[4]{n}}.$$

Here $\tilde{K}$ denotes a suitable positive constant, and $\mathcal{H}$ denotes the classical Hankel curve, *i.e.*, the noose-shaped curve that winds around 0 and starts and ends at $+\infty$ (*cf.* Figure 6).

Finally, using this result we get

$$[z^n]H_k\left(z, 1 + \frac{s}{\sqrt{n}}\right) = C(s)\rho_k\left(1 + \frac{s}{\sqrt{n}}\right)^{-n}n^{-5/4}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt[4]{n}}\right)\right), \quad \text{for } n \longrightarrow \infty,$$

with a constant $C(s)$ that depends on s. $\qquad\square$

Now we show that the characteristic function of our standardized sequence of random variables tends to the characteristic function of the normal distribution.

Lemma 3. *Let X_n be the total number of variables in a random lambda-term with at most k De Bruijn levels. Set $\sigma^2 := -\frac{\rho'_k(1)}{\rho_k(1)} - \frac{\rho''_k(1)}{\rho_k(1)} + \frac{\rho'_k(1)^2}{\rho_k(1)^2}$. If $\sigma^2 \neq 0$, then*

$$Z_n = \frac{X_n - \mathbb{E}X_n}{\sqrt{n}} \longrightarrow \mathcal{N}\left(0, \sigma^2\right).$$

Proof. For the standardised sequence of random variables Z_n we have with $\mu := \frac{\mathbb{E}X_n}{n}$

$$Z_n = \frac{X_n - \mathbb{E}X_n}{\sqrt{n}} = \frac{X_n}{\sqrt{n}} - \mu\sqrt{n}.$$

Its characteristic function reads as

$$\phi_{Z_n}(s) = \mathbb{E}(e^{isZ_n}) = e^{-is\mu\sqrt{n}}\phi_{X_n}\left(\frac{s}{\sqrt{n}}\right) = e^{-is\mu\sqrt{n}}\mathbb{E}(e^{\frac{isX_n}{\sqrt{n}}}) = \tag{14}$$

$$= e^{-is\mu\sqrt{n}}\frac{[z^n]H_k(z, e^{\frac{is}{\sqrt{n}}})}{[z^n]H_k(z,1)}. \tag{15}$$

From Proposition 3 we know

$$\frac{[z^n]H_k(z, 1+\frac{s}{\sqrt{n}})}{[z^n]H_k(z,1)} \sim C(s)\left(\frac{\rho_k(1+\frac{s}{\sqrt{n}})}{\rho_k(1)}\right)^{-n},$$

where the constant $C(s) \sim 1$ for $n \longrightarrow \infty$.

Thus,

$$\phi_{Z_n}(s) = e^{-is\mu\sqrt{n}}\frac{[z^n]H_k(z, e^{\frac{is}{\sqrt{n}}})}{[z^n]H_k(z,1)} \sim e^{-is\mu\sqrt{n}}\left(\frac{\rho_k\left(1+\frac{si}{\sqrt{n}}-\frac{s^2}{2n}+\mathcal{O}\left(\frac{|s^3|}{n^{3/2}}\right)\right)}{\rho_k(1)}\right)^{-n} =$$

$$= e^{-is\mu\sqrt{n}}\exp\left(-n\cdot\left(\log\left(1+\frac{\rho_k'is}{\rho_k\sqrt{n}}-\frac{s^2}{2n}\frac{\rho_k'}{\rho_k}+\frac{s^2}{2n}\frac{\rho_k''}{\rho_k}\right)+\mathcal{O}\left(\frac{|s^3|}{n^{3/2}}\right)\right)\right)$$

$$\sim e^{-is\mu\sqrt{n}}e^{-is\sqrt{n}\frac{\rho_k'}{\rho_k}}e^{\frac{s^2}{2}\left(\frac{\rho_k'}{\rho_k}+\frac{\rho_k''}{\rho_k}-\frac{\rho_k'^2}{\rho_k^2}\right)}.$$

Since we know that the expected value of the standardised random variable is zero, we get $\mu = -\frac{\rho_k'(1)}{\rho_k(1)} + o\left(\frac{1}{\sqrt{n}}\right)$, and thus

$$\phi_{Z_n}(s) \sim e^{-\frac{s^2\sigma^2}{2}},$$

with $\sigma^2 = -\frac{\rho_k'(1)}{\rho_k(1)} - \frac{\rho_k''(1)}{\rho_k(1)} + \frac{\rho_k'(1)^2}{\rho_k(1)^2}$, which completes the proof. $\qquad\square$

Thus, we get that the total number of leaves in lambda-terms with a bounded number of De Bruijn levels is asymptotically normally distributed.

5. Unary profile of random lambda-terms with bounded number of De Bruijn levels

5.1. **Leaves.** The aim of this section is the investigation of the distribution of the number of leaves in the different De Bruijn levels in closed lambda-terms with bounded number of De Bruijn levels. In order to do so, observe that each De Bruijn level in such a lambda-term corresponds to one or more binary trees that contain different types of leaves, where the maximal number of types corresponds to the respective level (*cf.* Figure 3), *i.e.*, in the i-th De Bruijn level there may be i different types of leaves. Let $\mathcal{C}$ be the class of binary trees. Using the notation from the previous sections we can specify this class by

$$\mathcal{C} = \mathcal{Z} + (\mathcal{A} \times \mathcal{C} \times \mathcal{C}).$$

Translating into bivariate generating functions $C(z, u)$ with z marking the size (*i.e.*, the total number of nodes) and u marking the number of leaves, yields $C(z, u) = \frac{1-\sqrt{1-4uz^2}}{2z}$.

Let $_{k-l}\tilde{H}_k(z, u)$ be the generating function of closed lambda-terms with at most k De Bruijn levels, where z marks the size and u marks the number of leaves on the $(k-l)$-th unary level ($0 \leq l \leq k$). The corresponding structures can be seen as $k+1$ nested binary trees, matching the $k+1$ De Bruijn levels. Each such level is a binary tree where some of its leaves are replaced by a unary node with a binary tree (belonging to the next level) attached to it. Thus, in the associated generating function, $C(z, u)$, the variable marking the leaves is replaced by $m + C(z, \ldots)$, where m is the number of possible De Bruijn indices (*i.e.* types of leaves) and $C(z, \ldots)$ the generating function of the binary trees in the next level. Altogether, we have

$$_{k-l}\tilde{H}_k(z, u) = C(z, C(z, 1 + \ldots + C(z, (k-l)\cdot u + \ldots + C(z, (k-1) + C(z, k)))\ldots)\ldots)).$$

This can be written as

$$_{k-l}\tilde{H}_k(z,u) = \frac{1 - \sqrt{\tilde{R}_{k+1,k}(z,u)}}{2z},$$

with

$$\tilde{R}_{1,k}(z,u) = 1 - 4z^2 k,$$

$$\tilde{R}_{i,k}(z,u) = 1 - 4z^2(k-i+1) - 2z + 2z\sqrt{\tilde{R}_{i-1,k}(z,u)}, \quad \text{for } 2 \le i \le k+1, i \ne l+1,$$

and

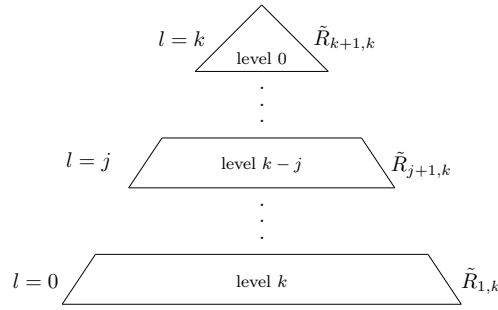$$\tilde{R}_{l+1,k}(z,u) = 1 - 4z^2 u(k-l) - 2z + 2z\sqrt{\tilde{R}_{l-1,k}(z,u)}.$$

FIGURE 7. A schematic sketch of a lambda-term with at most k De Bruijn levels that exemplifies the notation that is used within this section: If we investigate the number of leaves in the $(k-l)$-th De Bruijn level, for $0 \le l \le k$, a factor u is inserted in the recursive definition of the $(l+1)$-th radicand.

Remark. Note that the radicands $\tilde{R}_{i,k}$ that are introduced above are very similar to the radicands $R_{i,k}$ that were used in the previous section. The only difference is that now we have a u only in the $(l+1)$-th radicand, while in the previous case u was occurring in all radicands. Thus, from now on we will have further distinctions of cases now depending on the relative position (w.r.t. l) of the radicand(s) where the dominant singularity comes from.

This subsection consists of two subsubsections. In the first part we will derive the mean values for the number of leaves in the different De Bruijn levels and the second part deals with the distributions of the number of leaves in these levels.

5.1.1. *Mean values.* Now we want to determine the mean for the number of leaves in the different De Bruijn levels, *i.e.*

$$\mathbb{E}X_n = \frac{[z^n]\left(\frac{\partial}{\partial u} \, _{k-l}\tilde{H}_k(z,u)\right)\Big|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)},$$

where X_n denotes the number of leaves in the $(k-l)$-th De Bruijn level of a random closed lambda-term of size n with at most k De Bruijn levels.

In order to do so, we make the following considerations:

- $\frac{\partial}{\partial u}\tilde{R}_{i,k}(z,u) = 0 \quad \forall i < l+1$

- $\frac{\partial}{\partial u}\tilde{R}_{l+1,k}(z,u) = -4z^2(k-l)$

- $\frac{\partial}{\partial u}\tilde{R}_{i,k}(z,u) = z \cdot \frac{\frac{\partial}{\partial u}\tilde{R}_{i-1,k}(z,u)}{\sqrt{\tilde{R}_{i-1,k}(z,u)}} \quad \forall i > l+1$

Therefore we get

$$\left(\frac{\partial}{\partial u}\,_{k-l}\tilde{H}_k(z,u)\right)\bigg|_{u=1} = z^{k-l+1}(k-l)\prod_{i=l+1}^{k+1}\frac{1}{\sqrt{\tilde{R}_{i,k}(z,1)}}. \tag{16}$$

Again we perform a distinction of cases starting with k not being an element of the sequence $(N_j)_{j\in\mathbb{N}}$.

The case: $N_j < k < N_{j+1}$. Let $\tilde{\rho}_k(u)$ be the dominant singularity of $_{k-l}\tilde{H}_k(z,u)$, which we know comes from the $(j+1)$-th radicand $\tilde{R}_{j+1,k}(z,u)$. Obviously, $\tilde{\rho}_k(1) = \rho_k(1)$. Therefore we will again use the abbreviation $\rho_k := \tilde{\rho}_k(1)$.

From Proposition 2 we get the following expansions of the radicands for $u = 1$ and $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C}\setminus\mathbb{R}^-$:

- $\forall i < j+1$ (inner radicands) : $\tilde{R}_{i,k}(\rho_k - \epsilon, 1) = \tilde{R}_{i,k}(\rho_k, 1) + \mathcal{O}(|\epsilon|)$,

- $\tilde{R}_{j+1,k}(\rho_k - \epsilon, 1) = \tilde{\gamma}_{j+1}\epsilon + \mathcal{O}(|\epsilon|^2)$, with $\tilde{\gamma}_{j+1} = -\frac{\partial}{\partial z}\tilde{R}_{j+1,k}(\rho_k, 1)$,

- $\forall i > j+1$ (outer radicands) : $\tilde{R}_{i,k}(\rho_k - \epsilon, 1) = \tilde{a}_i + \tilde{b}_i\sqrt{\epsilon} + \mathcal{O}(|\epsilon|)$,

with

$$\tilde{a}_{i+1} = 1 - 4(k-i)\rho_k^2 - 2\rho_k + 2\rho_k\sqrt{\tilde{a}_i}, \tag{17}$$

and

$$\tilde{b}_{i+1} = \frac{\tilde{b}_i\rho_k}{\sqrt{\tilde{a}_i}} \quad \text{for} \quad j+2 \le i \le k, \tag{18}$$

where

$$\tilde{a}_{j+2} = 1 - 4(k-j-1)\rho_k^2 - 2\rho_k \text{ and } \tilde{b}_{j+2} = 2\rho_k\sqrt{\tilde{\gamma}_{j+1}}. \tag{19}$$

Thus, we have

- $\forall i < j+1$ (inner radicands) : $\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k-\epsilon,1)}} = \frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}} + \mathcal{O}(|\epsilon|)$,

- $\frac{1}{\sqrt{\tilde{R}_{j+1,k}(\rho_1-\epsilon,1)}} = \frac{1}{\sqrt{\tilde{\gamma}_{j+1}}}\epsilon^{-\frac{1}{2}} + \mathcal{O}(|\epsilon|^{\frac{1}{2}})$,

- $\forall i > j+1$ (outer radicands) : $\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k-\epsilon,1)}} = \frac{1}{\sqrt{\tilde{a}_i}} - \frac{\tilde{b}_i}{2\sqrt{\tilde{a}_i^3}}\epsilon^{\frac{1}{2}} + \mathcal{O}(|\epsilon|^{\frac{3}{2}})$.

Now we have to perform a distinction of cases whether the De Bruijn level that we are focusing on is below the $(k-j)$-th level or not (*i.e.*, whether l is below j or not).

<u>First case: $l > j$.</u> First let us remember that $l > j$ implies that the u is inserted in a radicand that is located outside the $(j+1)$-th. From (16) we get for $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C}\setminus\mathbb{R}^-$

$$\left(\frac{\partial}{\partial u}\,_{k-l}\tilde{H}_k(\rho_k - \epsilon, u)\right)\bigg|_{u=1} =$$

$$\rho_k^{k-l+1}(k-l)\left(\prod_{i=l+1}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}} - \sum_{m=l+1}^{k+1}\left(\frac{\tilde{b}_m}{2\sqrt{\tilde{a}_m^3}}\prod_{i=l+1,i\neq m}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}}\right)\epsilon^{\frac{1}{2}} + \mathcal{O}(|\epsilon|)\right).$$

By denoting the sum in the equation above with $\tilde{\delta}_l$ we can determine the coefficient of z^n by

$$[z^n]\left(\frac{\partial}{\partial u}\,_{k-l}\tilde{H}_k(z,u)\right)\bigg|_{u=1} = -\rho_k^{k-l+1}(k-l)\tilde{\delta}_l\left(\frac{1}{\rho_k}\right)^n\frac{n^{-\frac{3}{2}}}{\Gamma(-\frac{1}{2})}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right), \qquad \text{as } n \longrightarrow \infty,$$

and by using the asymptotics of the n-th coefficient of $_{k-l}\tilde{H}_k(z,1) = H_k(z,1)$ (see Theorem 7) we finally get for the mean asymptotically as $n \longrightarrow \infty$

$$\frac{[z^n]\left(\frac{\partial}{\partial u}\,_{k-l}\tilde{H}_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)} = \frac{-\rho_k^{k-l+1}(k-l)\tilde{\delta}_l}{h_k}\left(1 + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)\right).$$

Thus, we showed that there is only a small number of leaves in the De Bruijn levels below the $(k - j)$-th level. More precisely, the asymptotic mean of the number of leaves is $O(1)$ for all these lower levels.

Remark. The constant h_k can be expressed by means of the sequences $\tilde{a}_i$ and $\tilde{b}_i$ defined in Eqs. (17)-(19), thereby enabling a representation of the constant $C_{k,l} := \frac{-\rho_k^{k-l+1}(k-l)\tilde{\delta}_l}{h_k}$ that reads as

$$C_{k,l} = \frac{2(k-l)\rho_k^2}{a_{l+1}}\left(1 + \frac{\rho_k a_{l+1}}{a_{l+2}\sqrt{a_{l+1}}} + \frac{\rho_k^2 a_{l+1}}{a_{l+3}\sqrt{a_{l+2}}\sqrt{a_{l+1}}} + \dots\right). \tag{20}$$

This term can be used in order to investigate the asymptotic number of leaves in the lower De Bruijn levels more thoroughly.

Second case: $l \leq j$. Similar to the first case we get

$$\left(\frac{\partial}{\partial u}\, _{k-l}\tilde{H}_k(\rho_k - \epsilon, u)\right)\bigg|_{u=1} =$$

$$\rho_k^{k-l+1}(k-l)\left(\left(\prod_{i=l+1}^{j}\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}}\right)\left(\prod_{i=j+2}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}}\right)\frac{1}{\sqrt{\tilde{\gamma}_{j+1}}}\epsilon^{-\frac{1}{2}} + \text{ const. term} + \mathcal{O}(|\epsilon|^{\frac{1}{2}})\right).$$

By setting $\tilde{\phi}_{j+1,l} := \left(\prod_{i=l+1}^{j}\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}}\right)\left(\prod_{i=j+2}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}}\right)\frac{1}{\sqrt{\tilde{\gamma}_{j+1}}}$, we obtain for $n \longrightarrow \infty$

$$[z^n]\left(\frac{\partial}{\partial u}\, _{k-l}\tilde{H}_k(z, u)\right)\bigg|_{u=1} = \rho_k^{k-l+1}(k-l)\tilde{\phi}_{j+1,l}\left(\frac{1}{\rho_k}\right)^n\frac{n^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Thus, we get for the mean asymptotically as $n \longrightarrow \infty$

$$\frac{[z^n]\left(\frac{\partial}{\partial u}\, _{k-l}\tilde{H}_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)} = \frac{\rho_k^{k-l+1}(k-l)\Gamma(-\frac{1}{2})\tilde{\phi}_{j+1,l}}{\Gamma(\frac{1}{2})h_k} \cdot n\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right).$$

Hence, we proved that the asymptotic mean for the number of leaves in the De Bruijn levels above the $(k - j)$-th is $\Theta(n)$. So, altogether we can see that almost all of the leaves are located in the upper $j + 1$ De Bruijn levels.

The case: $k = N_j$. Now we will deal with the second case, where the bound k is an element of the sequence $(N_j)_{j \in \mathbb{N}}$.

We start by determining the expansions of the radicands around the dominant singularity $\tilde{\rho}_k(u)$ of $_{k-l}\tilde{H}_k(z, u)$ for $u = 1$ and $\epsilon \longrightarrow 0$ so that $\epsilon \in \mathbb{C} \setminus \mathbb{R}^-$ (cf. [8, Proposition 9]):

- $\forall i < j$ (inner radicands) : $\tilde{R}_{i,k}(\rho_k - \epsilon, 1) = \tilde{R}_{i,k}(\rho_k, 1) + \mathcal{O}(|\epsilon|)$,

- $\tilde{R}_{j,k}(\rho_k - \epsilon, 1) = \tilde{\gamma}_j\epsilon + \mathcal{O}(|\epsilon|^2)$ with $\tilde{\gamma}_j = -\frac{\partial}{\partial z}\tilde{R}_{j,k}(\rho_k, 1)$,

- $\tilde{R}_{j+1,k}(\rho_k - \epsilon, 1) = 2\tilde{\rho}_k\sqrt{\tilde{\gamma}_j}\epsilon^{\frac{1}{2}} + \mathcal{O}(|\epsilon|)$,

- $\forall i > j + 1$ (outer radicands) : $\tilde{R}_{i,k}(\rho_k - \epsilon, 1) = \tilde{a}_i + \tilde{b}_i\epsilon^{\frac{1}{4}} + \mathcal{O}(|\epsilon|)$,

  with $\tilde{a}_{i+1} = 1 - 4(k-i)\rho_k^2 - 2\rho_k + 2\rho_k\sqrt{\tilde{a}_i}$, and $\tilde{b}_{i+1} = \frac{\tilde{b}_i\rho_k}{\sqrt{\tilde{a}_i}}$ for $j + 2 \leq i \leq k$, with $\tilde{a}_{j+2} = 1 - 4(k-j)\rho_k^2 - 2\rho_k$ and $\tilde{b}_{j+2} = 2\rho_k\sqrt{\tilde{\gamma}_j}$.

Thus, we get

- $\forall i < j$ (inner radicands) : $\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k-\epsilon,1)}} = \frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}} + \mathcal{O}(|\epsilon|)$,

- $\frac{1}{\sqrt{\tilde{R}_{j,k}(\rho_k-\epsilon,1)}} = \frac{1}{\sqrt{\tilde{\gamma}_j}}\epsilon^{-\frac{1}{2}} + \mathcal{O}(|\epsilon|^{\frac{1}{2}})$,

- $\frac{1}{\sqrt{\tilde{R}_{j+1,k}(\tilde{\rho}_k-\epsilon,1)}} = \frac{1}{\sqrt{2\rho_k}\sqrt[4]{\tilde{\gamma}_j}}\epsilon^{-\frac{1}{4}} + \mathcal{O}(|\epsilon|^{\frac{1}{4}})$,

- $\forall i > j+1$ (outer radicands) : $\frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k - \epsilon, 1)}} = \frac{1}{\sqrt{\tilde{a}_i}} - \frac{\tilde{b}_i}{2\sqrt{\tilde{a}_i^3}}\epsilon^{\frac{1}{4}} + \mathcal{O}(|\epsilon|).$

We proceed analogously to the case where $N_j < k < N_{j+1}$, with the only difference that we have to distinguish between three cases now and since for $u = 1$ the j-th and the $(j+1)$-th radicand vanish simultaneously, we get a closed formula for the dominant singularity $\rho_k = \frac{1}{1+\sqrt{1+4(k-j)}}$.

<u>First case:</u> $l > j$. Let us again remember that $l > j$ implies that the u is inserted in the p-th radicand with $p > j+1$. From (16) we get for $\epsilon \in \mathbb{C} \setminus \mathbb{R}^-$ with $|\epsilon| \longrightarrow 0$

$$\left(\frac{\partial}{\partial u}\ _{k-l}\tilde{H}_k(\rho_k - \epsilon, u)\right)\bigg|_{u=1} = \rho_k^{k-l+1}(k-l)\prod_{i=l+1}^{k+1}\left(\frac{1}{\sqrt{\tilde{a}_i}} - \frac{\tilde{b}_i}{2\sqrt{\tilde{a}_i^3}}\epsilon^{\frac{1}{4}} + \mathcal{O}(|\epsilon|)\right) =$$

$$\left(\frac{1}{1+\sqrt{1+4(k-j)}}\right)^{k-l+1}(k-l)\left(\prod_{i=l+1}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}} - \sum_{m=l+1}^{k+1}\left(\frac{\tilde{b}_m}{2\sqrt{\tilde{a}_m^3}}\prod_{i=l+1, i\neq m}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}}\right)\epsilon^{\frac{1}{4}} + \mathcal{O}(|\epsilon|^{\frac{1}{2}})\right).$$

As in the previous case, we set $\tilde{\delta}_l := \sum_{m=l+1}^{k+1}\left(\frac{\tilde{b}_m}{2\sqrt{\tilde{a}_m^3}}\prod_{\substack{i=l+1 \\ i\neq m}}^{k+1}\frac{1}{\sqrt{\tilde{a}_i}}\right)$. Extracting the n-th coefficient and using the asymptotics of $[z^n]_{k-l}\tilde{H}_k(z,1) = [z^n]H_k(z,1) = \frac{-\rho_k^{k-j-1}b_{j+2}n^{-5/4}}{4\rho_k\Gamma(-1/4)\prod_{i=j+2}^{k+1}\sqrt{a_i}}\rho_k^{-n}\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right)$ we have for $n \longrightarrow \infty$

$$\frac{[z^n]\left(\frac{\partial}{\partial u}\ _{k-l}\tilde{H}_k(z, u)\right)|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)} = \frac{-4\rho_k^{j-l+3}(k-l)\tilde{\delta}_l \prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}\left(1+\mathcal{O}\left(n^{-\frac{1}{4}}\right)\right).$$

Thus, as in the previous case ($k \in (N_j, N_{j+1})$) the asymptotic mean for the number of leaves in the De Bruijn levels below the $(k-j)$-th level is $O(1)$.

Furthermore the constant $D_{k,l} := \frac{-4\rho_k^{j-l+3}(1)(k-l)\tilde{\delta}_l \prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}$ can be simplified to

$$D_{k,l} = \frac{k-l}{2\lambda_{l-j}}\left(1 + \frac{\sqrt{\lambda_{l-j}}}{2\lambda_{l-j+1}} + \frac{\sqrt{\lambda_{l-j}}}{4\lambda_{l-j+2}\sqrt{\lambda_{l-j+1}}} + \frac{\sqrt{\lambda_{l-j}}}{8\lambda_{l-j+3}\sqrt{\lambda_{l-j+2}}\sqrt{\lambda_{l-j+1}}} + \ldots\right), \quad (21)$$

with the sequence λ_i defined by $\lambda_0 = 0$ and $\lambda_{i+1} = i + 1 + \sqrt{\lambda_i}$ for $i \geq 0$.

<u>Second case:</u> $l = j$. Thus, the u is inserted in the $(j+1)$-th radicand. In this case we get

$$\frac{[z^n]\left(\frac{\partial}{\partial u}\ _{k-l}\tilde{H}_k(z, u)\right)|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)} = \frac{-4\rho_k^3(k-j)\Gamma(-1/4)\psi_j \prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(\frac{1}{4})b_{j+2}} \cdot \sqrt{n}\ \left(1+\mathcal{O}\left(n^{-\frac{1}{4}}\right)\right), \quad (22)$$

with

$$\psi_j = \frac{1}{\sqrt{2\rho_k}\sqrt[4]{\tilde{\gamma}_j}}\prod_{i=j+2}^{k+1}\frac{1}{\sqrt{a_i}}. \quad (23)$$

The constant $\hat{D}_{k,l} := \frac{-4\rho_k^3(k-j)\Gamma(-1/4)\psi_j \prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(\frac{1}{4})b_{j+2}}$ simplifies to

$$\hat{D}_{k,l} = \frac{-\Gamma(-1/4)(k-j)\sqrt{\rho_k}}{\Gamma(1/4)\sqrt{\tilde{\gamma}_j}}.$$

In order to get some information on the magnitude of this factor we would have to investigate $\tilde{\gamma}_j = -\frac{\partial}{\partial z}\tilde{R}_{j,k}(\rho_k, 1)$, which seems to get rather involved. However, taking a look at Equation (22) we can see that there are already considerably more unary nodes in the $(k-j)$-th De Bruijn level, namely $\Theta(\sqrt{n})$.

<u>Third case:</u> $l < j$. The third case gives for $n \longrightarrow \infty$

$$\frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}\tilde{H}_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}\tilde{H}_k(z,1)} = \frac{-4\rho_k^{j-l+3}(k-l)\Gamma(-1/4)\chi_j \prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(3/4)b_{j+2}} n \left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

with

$$\chi_j = \frac{1}{\sqrt{\tilde{\gamma}_j}}\psi_j \left(\prod_{i=l+1}^{j-1} \frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}}\right),$$

where ψ_j is defined as in (23).

Thus, we proved that asymptotically there is an average of $\Theta(n)$ leaves in the upper j De Bruijn levels. The constant $\tilde{D}_{k,l} := \frac{-4\rho_k^{j-l+3}(k-l)\Gamma(-1/4)\chi_j \prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(3/4)b_{j+2}}$ can be rewritten as

$$\tilde{D}_{k,l} = \frac{-\Gamma(-1/4)(k-l)\rho_k^{j-l}}{\Gamma(3/4)\tilde{\gamma}_j} \prod_{i=l+1}^{j-1} \frac{1}{\sqrt{\tilde{R}_{i,k}(\rho_k,1)}}.$$

The following proposition sums up all the results that we obtained within this section.

Proposition 4. *Let X_n denote the number of leaves in the $(k-l)$-th De Bruijn level in a random lambda-term of size n with at most k De Bruijn levels.*

If $k \in (N_j, N_{j+1})$, then we get for the asymptotic mean when $n \longrightarrow \infty$

- *in the case $l > j$:*

$$\mathbb{E}X_n = \frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}H_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}H_k(z,1)} = C_{k,l}\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

- *and in the case $l \leq j$:*

$$\mathbb{E}X_n = \frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}H_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}H_k(z,1)} = \tilde{C}_{k,l} \cdot n \left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

with constants $C_{k,l}$ and $\tilde{C}_{k,l}$ depending on l and k.

If $k = N_j$, then the asymptotic mean for $n \longrightarrow \infty$ reads as

- *in the case $l > j$:*

$$\mathbb{E}X_n = \frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}H_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}H_k(z,1)} = D_{k,l}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

- *in the case $l = j$:*

$$\mathbb{E}X_n = \frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}H_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}H_k(z,1)} = \hat{D}_{k,l} \cdot \sqrt{n} \left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

- *and in the case $l < j$:*

$$\mathbb{E}X_n = \frac{[z^n] \left(\frac{\partial}{\partial u}\ _{k-l}H_k(z,u)\right)|_{u=1}}{[z^n]_{k-l}H_k(z,1)} = \tilde{D}_{k,l} \cdot n \left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

with constants $D_{k,l}$, $\hat{D}_{k,l}$ and $\tilde{D}_{k,l}$ depending on l and k.

All the constants occurring in Proposition 4 have been calculated explicitly and can be obtained for every fixed k. In particular, we investigated $D_{k,l}$ in order to show that for large k the number of leaves in the De Bruijn levels that are closer to the root is smaller (*cf.* Figure 8). In fact, they rapidly tend to zero for k tending to infinity.

Proposition 5. *Let us consider a random closed lambda-term of size n with at most k De Bruijn levels and let us consider the case $k = N_j$. Then the average number of leaves in De Bruijn level L, with $0 \leq L \leq k - j - 1$, is asymptotically equal to $D_{k,k-L}$, defined in (21). It behaves like*

$$D_{k,k-L} \sim \frac{L}{2(k-j-L)} \qquad as\ k \longrightarrow \infty.$$

Proof. The proposition follows directly by investigating this constant $D_{k,l}$. The asymptotics for the sequence λ_i (defined by $\lambda_0 = 0$ and $\lambda_{i+1} = i+1+\sqrt{\lambda_i}$ for $i \geq 0$) can be obtained by bootstrapping (see [20]). We obtain $\lambda_i \sim i$, as $i \longrightarrow \infty$. □

Remark. Note that the expression for $D_{k,l}$ (*cf.* Equ. (21)) can be obtained by plugging $\tilde{a}_{j+l} = 4\rho_k^2 \lambda_{l-1}$ into the equation for $C_{k,l}$ (*cf.* Equ. (20)). However, this relation is solely valid for the case $k = N_j$ and thus, Proposition 5 holds just for the constants $D_{k,l}$. Nonetheless, we expect that by means of some suitable estimates for the $\tilde{a}_i's$ one can obtain a similar behaviour for the constants $C_{k,l}$. Since computations get rather involved, we omitted any further investigations of these constants within this paper. Anyway, we can conclude that in both cases, whether k is an element of $(N_i)_{i>0}$ or not, a random closed lambda-term with at most k De Bruijn levels has very few leaves in its lowest levels if k is large.

5.1.2. *Distributions.* Now that we derived the mean values for the number of leaves in the different De Bruijn levels, we are interested in their distribution. Therefore we distinguish again between the cases of k being an element of the sequence $(N_i)_{i\geq 0}$ or not.

The case: $N_j < k < N_{j+1}$. We know that the generating function $_{k-l}\tilde{H}_k(z,u)$ consists of $k+1$ nested radicals, where a u is inserted in the $(l+1)$-th radicand counted from the innermost one. Additionally we know that for $N_j < k < N_{j+1}$ the dominant singularity $\tilde{\rho}_k(u)$ comes from the $(j+1)$-th radicand. Therefore, for $l > j$ the function $\tilde{\rho}_k(u)$ is independent of u, which is the reason why we do not get a quasi-power in that case. Thus, for the first $k-j$ levels of the lambda-DAG (*i.e.* the case $l > j$), where there are just a few leaves, we can not say something about the distribution of the leaves so far. It might be a degenerated distribution.

However, in case that $l \leq j$ (*i.e.*, for the upper levels where there are a lot of leaves) we will use the Quasi-Powers Theorem to show that the number of leaves in the $(k-j)$-th until the k-th level is asymptotically normally distributed.

Analogously as we did in Section 4.1 we can show that

$$\frac{[z^n] \, _{k-l}\tilde{H}_k(z,u)}{[z^n] \, _{k-l}\tilde{H}_k(z,1)} = \frac{\tilde{h}_k(u)}{h_k}\left(\frac{\rho_k}{\tilde{\rho}_k(u)}\right)^n\left(1+\mathcal{O}\left(\frac{1}{n}\right)\right). \tag{24}$$

We can easily see that Equation (24) has the desired shape for the Quasi-Powers Theorem. Hence, assuming that $\tilde{B}''(1)+\tilde{B}'(1)-\tilde{B}'(1)^2 \neq 0$, where $\tilde{B}(u) = \frac{\rho_k}{\tilde{\rho}_k(u)}$, the Quasi-Powers Theorem can be applied, which proves that the number of leaves in a De Bruijn level that is above the $(k-j-1)$-th level is asymptotically normally distributed.

The case: $k = N_j$. As is the previous case we do not know the distribution of the number of leaves in the lowest $k-j$ De Bruijn levels (*i.e.*, the levels 0 to $k-j-1$), due to the fact that for these levels the function $\tilde{\rho}_k(u)$ does not depend on u. It might also be a degenerated distribution.

In Section 4.2 we showed that the dominant singularity comes from the j-th radicand when u is in a neighbourhood of 1. Thus, for the case that $l = j$, where we insert a u in the $j+1$-th radicand, the dominant singularity $\rho_k(u)$ does still do not depend on u. Therefore we also do not know the distribution of the leaves in the $(k-j)$-th De Bruijn level. It seems very unlikely that the number of leaves in this level will be asymptotically normally distributed, but further studies on this subject might be very interesting.

Now we are going to show that the number of leaves in the upper j De Bruijn levels (*i.e.*, from the $(k-j+1)$-th to the k-th level) is asymptotically normally distributed. In order to do so we proceed analogously as in Section 4.2 for the total number of leaves. Therefore for $l < j$ we set again $z = \tilde{\rho}_k(u)(1+\frac{t}{n})$ and $u = 1+\frac{s}{n}$ and obtain expansions that behave just as the ones in Lemma 2. The only differences that occur concern the constants and therefore do not alter our results for the normal distribution.

Thus, Theorem 3 is proved. Figure 8 summarizes the results that we obtained in Section 5.1 and illustrates a combinatorial interpretation of the occurring phenomena.
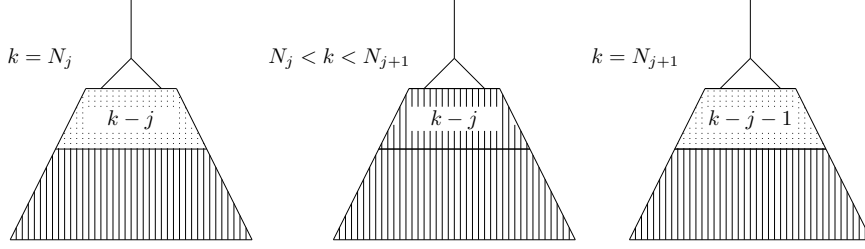
FIGURE 8. (1) In the $(k - j)$-th Be Bruijn level $(l = j)$ are considerably more leaves than in the lower levels, but still less leaves then in the levels above. (2) With growing k the $(k - j)$-th Be Bruijn level gets filled with leaves, while the number of leaves in the next level below (*i.e.*, the $(k - j - 1)$-th) slowly increases. (3) As soon as k reaches the next element of the sequence $(N_j)_{j \geq 0}$, namely $k = N_{j+1}$ the $(k - j - 1)$-th De Bruijn level immediately contains considerably more leaves than the levels below.

5.2. **Unary nodes.** Now we want to investigate the number of unary nodes among the different De Bruijn levels. Let $C(z, u)$ again denote the generating function of the class of binary trees where z marks the total number of nodes and u marks the number of leaves, *i.e.*, $C(z, u) = \frac{1 - \sqrt{1 - 4z^2 u}}{2z}$. The bivariate generating function $_{k-l}\bar{H}_k(z, w)$ of the class of closed lambda-terms with at most k De Bruijn levels, where z marks the size and w the number of unary nodes in the $(k - l)$-th De Bruijn level, can then be expressed in terms of $C(z, u)$ by

$$C(z, C(z, 1 + \ldots + C(z, (k - l) + w \cdot C(z, \ldots (k - 1) + C(z, k)) \ldots) \ldots)).$$

This can be rewritten to

$$_{k-l}\bar{H}_k(z, w) = \frac{1 - \sqrt{\bar{R}_{k+1,k}(z, w)}}{2z},$$

with

$$\bar{R}_{1,k}(z, w) = 1 - 4z^2 k,$$

$$\bar{R}_{i,k}(z, w) = 1 - 4z^2(k - i + 1) - 2z + 2z\sqrt{\bar{R}_{i-1,k}(z, w)} \quad \text{for } 2 \leq i \leq k + 1, \quad i \neq l + 1,$$

$$\bar{R}_{l+1,k}(z, w) = 1 - 4z^2(k - l) - 2zw + 2zw\sqrt{\bar{R}_{l,k}(z, w)}.$$

Thus, for the derivatives we get

$$\frac{\partial \bar{R}_{i,k}(z, w)}{\partial w} = 0 \quad \text{for } i < l + 1,$$

$$\frac{\partial \bar{R}_{l+1,k}(z, w)}{\partial w} = -2z + 2z\sqrt{\bar{R}_{l,k}(z, w)},$$

$$\frac{\partial \bar{R}_{i,k}(z, w)}{\partial w} = \frac{z}{\sqrt{\bar{R}_{i-1,k}}} \frac{\partial \bar{R}_{i-1,k}(z, w)}{\partial w} \quad \text{for } i > l + 1,$$

which implies

$$\frac{\partial_{k-l}\bar{H}_k(z, w)}{\partial w}\bigg|_{w=1} = \frac{z^{k-l}}{2} \prod_{i=l+1}^{k+1} \frac{1}{\sqrt{\bar{R}_{i,k}}} \left(1 - \sqrt{\bar{R}_{l,k}}\right).$$

As in the previous section we distinguish between different cases.

5.2.1. **The case:** $k = N_j$.

First case: $l > j + 1$. Inserting the expansion of the radicands $\bar{R}_{i,k}$ and simplifying yields for $n \longrightarrow \infty$

$$[z^n]_{k-l}\bar{H}_k(z,w)\Big|_{w=1} = \rho_k^{k-l}2\alpha_l \frac{n^{-5/4}}{\Gamma(-1/4)}\rho_k^{-n}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

with

$$\alpha_l = -\frac{b_l}{2\sqrt{a_l}}\prod_{i=l+1}^{k+1}\frac{1}{\sqrt{a_i}} - \sum_{m=l+1}^{k+1}\frac{b_m}{2\sqrt{a_m^3}}\prod_{\substack{i=l+1\\i\neq m}}^{k+1}\frac{1}{\sqrt{a_i}}(1 - \sqrt{a_l}), \tag{25}$$

where $a_i := \tilde{a}_i = a_i(1)$ and $b_i := \tilde{b}_i = b_i(1)$ are defined in the previous sections and result from the expansions of the radicands.

Thus, in this case the expected value of the number of unary nodes in the $(k-l)$-th De Bruijn level reads as

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\alpha_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right), \quad \text{as } n \longrightarrow \infty.$$

Furthermore, the constant $\frac{-2\alpha_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}$ can be simplified to

$$1 + \left(\frac{1}{4\rho_k\lambda_{l-j}} - \frac{\sqrt{\lambda_{l-j-1}}}{2\lambda_{l-j}}\right)\left(1 + \frac{\lambda_{l-j}}{2\lambda_{l-j+1}\sqrt{\lambda_{l-j}}} + \frac{\lambda_{l-j}}{2^2\lambda_{l-j+2}\sqrt{\lambda_{l-j+1}}\sqrt{\lambda_{l-j}}} + \dots\right),$$

with the sequence λ_i defined by $\lambda_0 = 0$ and $\lambda_{i+1} = i + 1 + \sqrt{\lambda_i}$ for $i \geq 0$.

Since the second summand is almost zero for l being close to k and large k, this implies that the number of unary nodes in these levels (close to the root) is close to one for large k.

Second case: $l = j + 1$. For $n \longrightarrow \infty$ we get

$$[z^n]_{k-l}\bar{H}_k(z,w)\Big|_{w=1} = \rho_k^{k-l}2\zeta_l\frac{n^{-5/4}}{\Gamma(-1/4)}\rho_k^{-n}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

with

$$\zeta_l = -\sqrt{2\rho_k}(\rho_k\tilde{\gamma})^{1/4}\prod_{i=j+2}^{k+1}\frac{1}{\sqrt{a_i}} - \sum_{m=j+2}^{k+1}\frac{b_m}{2\sqrt{a_m^3}}\prod_{\substack{i=j+2\\i\neq m}}^{k+1}\frac{1}{\sqrt{a_i}}.$$

Thus,

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\zeta_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right), \quad \text{as } n \longrightarrow \infty.$$

In this case the constant $\frac{-2\zeta_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}}$ simplifies to

$$1 + \frac{1}{4\rho_k}\left(1 + \frac{1}{2\lambda_2} + \frac{1}{2^2\lambda_3\sqrt{\lambda_2}} + \dots\right).$$

So, the expected number of unary nodes in the $(k-j-1)$-th De Bruijn level behaves exactly like in the lower levels. Starting from the next level a change in the behaviour can be determined, as we will see in the following.

<u>Third case</u>: $l = j$. For $n \longrightarrow \infty$ we have

$$[z^n]_{k-l}\bar{H}_k(z,w)\Big|_{w=1} = \rho_k^{k-l} 2\beta_l \frac{n^{-3/4}}{\Gamma(1/4)} \rho_k^{-n} \left(1 + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right),$$

with

$$\beta_l = \frac{1}{\sqrt{2\rho_k}\sqrt[4]{\rho_k\tilde{\gamma}_j}} \prod_{i=j+2}^{k+1} \frac{1}{\sqrt{a_i}}.$$

Thus,

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\beta_l\rho_k^2\Gamma(-1/4)\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(1/4)b_{j+2}} \cdot \sqrt{n}\left(1 + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right), \qquad \text{as } n \longrightarrow \infty.$$

The constant $\frac{-2\beta_l\rho_k^2\Gamma(-1/4)\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(1/4)b_{j+2}}$ can be written as

$$\frac{-\Gamma(-1/4)}{2\Gamma(1/4)\sqrt{\rho_k\tilde{\gamma}_j}}.$$

The expected number of unary nodes in this "separating level" is therefore asymptotically $\Theta(\sqrt{n})$ (as was the number of leaves).

<u>Fourth case</u>: $l < j$. For $n \longrightarrow \infty$ we get

$$[z^n]_{k-l}\bar{H}_k(z,w)\Big|_{w=1} = \rho_k^{k-l} 2\epsilon_l \frac{n^{-1/4}}{\Gamma(3/4)} \rho_k^{-n} \left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),$$

with

$$\epsilon_l = \frac{1}{\sqrt{2\rho_k}\sqrt[4]{\rho_k\tilde{\gamma}_j}\sqrt{\rho_k\tilde{\gamma}_j}} \prod_{i=j+2}^{k+1}\frac{1}{\sqrt{a_i}} \prod_{m=l+1}^{j-1}\frac{1}{\sqrt{\tilde{R}_{m,k}}}\left(1 - \sqrt{\bar{R}_{l,k}}\right).$$

Thus,

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\epsilon_l\rho_k^{j-l+2}\Gamma(-1/4)\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(3/4)b_{j+2}} \cdot n\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right), \qquad \text{as } n \longrightarrow \infty.$$

The constant $\frac{-2\epsilon_l\rho_k^{j-l+2}\Gamma(-1/4)\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\Gamma(3/4)b_{j+2}}$ can be simplified to

$$\frac{-\rho_k^{j-l+1}\Gamma(-1/4)}{2\Gamma(3/4)\tilde{\gamma}_j} \prod_{m=l+1}^{j-1}\frac{1}{\sqrt{\tilde{R}_{m,k}}}\left(1 - \sqrt{\bar{R}_{l,k}}\right).$$

Hence, analogously to the number of leaves, we proved that the number of unary nodes on the upper $j + 1$ De Bruijn levels is $\Theta(n)$.

5.2.2. **The case:** $N_j < k < N_{j+1}$. This case works analogously to the previous one. Thus, we just give the results for the expected values.

<u>First case</u>: $l > j + 1$. In this case, the expected value is entirely equal to the mean for the case $k = N_j$ and $l > j + 1$. So, with α_l defined as in (25), we have for $n \longrightarrow \infty$

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\alpha_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}} \cdot n\left(1 + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right).$$

Second case: $l = j + 1$. In the second case, the constant differs a little bit, but the result stays qualitatively unaltered. We get

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\mu_l\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}}{b_{j+2}} \cdot n\left(1 + \mathcal{O}\left(n^{-\frac{1}{2}}\right)\right), \quad \text{as } n \longrightarrow \infty,$$

with

$$\mu_l = -\sum_{m=j+2}^{k+1}\frac{b_m}{2\sqrt{a_m^3}}\prod_{\substack{i=j+2\\i\neq m}}^{k+1}\frac{1}{\sqrt{a_i}} + \sqrt{\rho_k\tilde{\gamma}_{j+1}}\prod_{i=j+2}^{k+1}\frac{1}{\sqrt{a_i}}.$$

Third case: $l < j + 1$. For $n \longrightarrow \infty$ we have

$$\frac{[z^n]\frac{\partial}{\partial w}{}_{k-l}\bar{H}_k(z,w)}{[z^n]_{k-l}\bar{H}_k(z,1)} = \frac{-2\theta_l\Gamma(-1/2)\rho_k^{j-l+2}\prod_{m=j+2}^{k+1}\sqrt{a_m}\prod_{s=l+1}^{j}\sqrt{\bar{R}_s}}{b_{j+2}\Gamma(1/2)} \cdot n\left(1 + \mathcal{O}\left(\frac{1}{n}\right)\right),$$

with

$$\theta_l = \prod_{i=j+2}^{k+1}\frac{1}{\sqrt{a_i}}\frac{1}{\sqrt{\rho_k\tilde{\gamma}_{j+1}}}\left(1 - \sqrt{\bar{R}_{j,k}}\right).$$

Thus, the expected number of unary nodes in the last $j + 1$ De Bruijn levels is asymptotically $\Theta(n)$.

5.3. **Binary nodes.** In this section we want to calculate the mean values of the number of binary nodes in the different De Bruijn levels. We denote by $C(z, v, u)$ the generating function of the class of binary trees where z marks the total number of nodes, v marks the number of binary nodes, and u marks the number of leaves. Thus, we have

$$C(z, v, u) = \frac{1 - \sqrt{1 - 4z^2uv}}{2zv}. \tag{26}$$

Using this generating function, we can write the bivariate generating function ${}_{k-l}H_k(z,v)$ of the class of closed lambda-terms with z marking the size, and v marking the the number of binary nodes on the $(k - l)$-th De Bruijn level as

$$C(z, 1, C(z, 1, 1 + C(z, 1, 2 + \ldots + C(z, v, (k - l) + \ldots + C(z, 1, k))\ldots)\ldots))). \tag{27}$$

Plugging in Equation (26) into (27) gives

$$_{k-l}\breve{H}_k(z, v) = \frac{1 - \sqrt{\breve{R}_{k+1,k}(z, v)}}{2z},$$

with

$$\breve{R}_{1,k}(z, v) = 1 - 4z^2k,$$

$$\breve{R}_{i,k}(z, v) = 1 - 4z^2(k - i + 1) - 2z + 2z\sqrt{R_{i-1,k}(z, v)}, \quad \text{for } 2 \leq i \leq k + 1, \ i \neq l + 1, \ i \neq l + 2,$$

$$\breve{R}_{l+1,k}(z, v) = 1 - 4z^2(k - l)v - 2zv + 2zv\sqrt{\breve{R}_{l,k}(z, v)},$$

$$\breve{R}_{l+2,k}(z, v) = 1 - 4z^2(k - l - 1) - \frac{2z}{v} + \frac{2z}{v}\sqrt{\breve{R}_{l+1,k}(z, v)}.$$

Thus, for the derivatives we get

$$
\frac{\partial \breve{R}_{i,k}(z,v)}{\partial v} = 0 \quad \text{for } i < l+1,
$$

$$
\frac{\partial \breve{R}_{l+1,k}(z,v)}{\partial v} = -4z^2(k-l) - 2z + 2z\sqrt{\breve{R}_{l,k}(z,v)},
$$

$$
\frac{\partial \breve{R}_{l+2,k}(z,v)}{\partial v} = \frac{2z}{v^2} - \frac{2z}{v^2}\sqrt{\breve{R}_{l+1,k}} + \frac{z}{v}\frac{1}{\sqrt{\breve{R}_{l+1,k}}}\frac{\partial \breve{R}_{l+1,k}(z,v)}{\partial v},
$$

$$
\frac{\partial \breve{R}_{i,k}(z,v)}{\partial v} = z\frac{1}{\sqrt{\breve{R}_{i-1,k}}}\frac{\partial \breve{R}_{i-1,k}(z,v)}{\partial v} \quad \text{for } i > l+2.
$$

Finally, we have

$$
\frac{\partial_{k-l}\breve{H}_k(z,v)}{\partial v}\bigg|_{v=1} = \prod_{i=l+2}^{k+1}\frac{1}{\sqrt{\breve{R}_{i,k}}}\left(-\frac{z^{k-l-1}}{2} + \frac{\sqrt{\breve{R}_{l+1,k}}z^{k-l-1}}{2} + \frac{z^{k-l}}{2\sqrt{\breve{R}_{l+1,k}}}\right.
$$
$$
\left. - \frac{z^{k-l}\sqrt{\breve{R}_{l,k}}}{2\sqrt{\breve{R}_{l+1,k}}} + \frac{z^{k-l+1}(k-l)}{\sqrt{\breve{R}_{l+1,k}}}\right). \tag{28}
$$

Analogously to the previous sections we have to distinguish between different cases. For the case $k = N_j$ and $l > j+1$ we get for $n \longrightarrow \infty$

$$
[z^n]_{k-l}\breve{H}_k(z,v) = \xi_l\frac{n^{-5/4}}{\Gamma(-1/4)}\rho_k^{-n}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right),
$$

with

$$
\xi_l = -\sum_{m=l+2}^{k+1}\frac{b_m}{2\sqrt{a_m^3}}\prod_{\substack{i=l+2\\i\neq m}}\frac{1}{\sqrt{a_i}}\left(\frac{\rho_k^{k-l-1}(\sqrt{a_{l+1}}-1)}{2} + \frac{\rho_k^{k-l+1}(k-l)}{\sqrt{a_{l+1}}} + \frac{\rho_k^{k-l}(1-\sqrt{a_l})}{2\sqrt{a_{l+1}}}\right)
$$
$$
+ \prod_{i=l+2}^{k+1}\frac{1}{\sqrt{a_i}}\left(\frac{\rho_k^{k-l-1}b_{l+1}}{2\sqrt{a_{l+1}}} - \frac{\rho_k^{k-l+1}(k-l)b_{l+1}}{2\sqrt{a_{l+1}^3}} + \frac{b_{l+1}}{2\sqrt{a_{l+1}^3}}(\sqrt{a_l}-1) - \frac{\rho_k^{k-l}b_l}{4\sqrt{a_l}\sqrt{a_{l+1}}}\right).
$$

Thus,

$$
\frac{[z^n]\frac{\partial}{\partial v}{}_{k-l}\breve{H}_k(z,v)}{[z^n]_{k-l}\breve{H}_k(z,1)} = \frac{-4\xi_l\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\rho_k^{k-j}b_{j+2}}\left(1 + \mathcal{O}\left(n^{-\frac{1}{4}}\right)\right), \quad \text{as } n \longrightarrow \infty.
$$

We performed a thorough investigation of the constant $\frac{-4\xi_l\prod_{m=j+2}^{k+1}\sqrt{a_m}}{\rho_k^{k-j}b_{j+2}}$ and showed that it is almost zero, in case l is close to $k+1$ and k is large,, i.e., if we consider a very low De Bruijn level, that is close to the root.

Due to Equation (28) calculations get rather involved. Since the methods that are used are the same as in the previous section, we will omit further calculations. However, the results resemble the ones that we got in Section 5.1 for the number of leaves. The only difference appears in the constants, but qualitatively also these constants behave equally.

6. Conclusion

Our investigation was triggered by the striking observation that the asymptotic number of lambda-DAGs with bounded number of De Bruijn levels, say bounded by k, and n vertices is of the form $\rho^n n^{-3/2}$ except if the bound k belongs to some peculiar doubly exponentially growing sequence. There was no apparent reason why bounding the number of De Bruijn levels by 8 is substantially different from setting the bound to 7 or 9.

The results in this paper showed that the vertices corresponding to the variables in the associated lambda-terms gather at the bottom of the lambda-DAG, meaning the De Bruijn levels of highest order within the lambda-DAG. Precisely, in each of the last ℓ_n levels, where $\ell_n = \Theta(\log \log k)$, we find $Theta(n)$ variables. The other levels contain only a bounded number of variables. As the bound grows, the higher levels become fuller and fuller and whenever k reaches a value that makes ℓ_n jump to the next integer, a further De Bruijn level becomes populated with variables. In this stage, there are only $\Theta(\sqrt{n})$ variables, but for the next value of k this level gets densely populated with variables, just as the other levels of high order. This shows that there is a structural difference within the classes of lambda-terms with at most k De Bruijn levels, depending on whether the bound belongs to $(N_i)_{i \geq 0}$ or not. The distribution of the variables, in particular the fact that a further level has to contain a larger but still fairly small number of variables apparently has some slight affects on the degrees of freedom to choose the bindings which modifies the subexponential term in the asymptotics.

References

[1] Henk P. Barendregt. *The Lambda Calculus, its Syntax and Semantics*, volume 40 of *Studies in Logic (London)*. College Publications, London, 2012. [Reprint of the 1984 revised edition, MR0774952], With addenda for the 6th imprinting, Mathematical Logic and Foundations.

[2] Maciej Bendkowski, Olivier Bodini, and Sergey Dovgal. Polynomial tuning of multiparametric combinatorial samplers. In *2018 Proceedings of the Fifteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 92–106. SIAM, Philadelphia, PA, 2018.

[3] Maciej Bendkowski, Katarzyna Grygiel, Pierre Lescanne, and Marek Zaionc. Combinatorics of λ-terms: a natural approach. *J. Logic Comput.*, 27(8):2611–2630, 2017.

[4] Maciej Bendkowski, Katarzyna Grygiel, and Paul Tarau. Boltzmann samplers for closed simply-typed lambda terms. In *Practical aspects of declarative languages*, volume 10137 of *Lecture Notes in Comput. Sci.*, pages 120–135. Springer, Cham, 2017.

[5] Maciej Bendkowski, Katarzyna Grygiel, and Marek Zaionc. On the likelihood of normalization in combinatory logic. *J. Logic Comput.*, 27(7):2251–2269, 2017.

[6] O. Bodini, D. Gardy, and A. Jacquot. Asymptotics and random sampling for BCI and BCK lambda terms. *Theoret. Comput. Sci.*, 502:227–238, 2013.

[7] Olivier Bodini, Danièle Gardy, and Bernhard Gittenberger. Lambda terms of bounded unary height. In *ANALCO11—Workshop on Analytic Algorithmics and Combinatorics*, pages 23–32. SIAM, Philadelphia, PA, 2011.

[8] Olivier Bodini, Danièle Gardy, Bernhard Gittenberger, and Zbigniew Gołębiewski. On the number of unary-binary tree-like structures with restrictions on the unary height. *Ann. Comb.*, 22(1):45–91, 2018.

[9] Olivier Bodini, Danièle Gardy, Bernhard Gittenberger, and Alice Jacquot. Enumeration of generalized *BCI* lambda-terms. *Electron. J. Combin.*, 20(4):Paper 30, 23, 2013.

[10] Olivier Bodini and Bernhard Gittenberger. On the asymptotic number of $BCK(2)$-terms. In *ANALCO14—Meeting on Analytic Algorithmics and Combinatorics*, pages 25–39. SIAM, Philadelphia, PA, 2014.

[11] Olivier Bodini, Bernhard Gittenberger, and Zbigniew Gołębiewski. Enumerating lambda terms by weighted length of their de Bruijn representation. *Discrete Appl. Math.*, 239:45–61, 2018.

[12] Olivier Bodini and Yann Ponty. Multi-dimensional Boltzmann sampling of languages. In *21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA'10)*, Discrete Math. Theor. Comput. Sci. Proc., AM, pages 49–63. Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2010.

[13] Olivier Bodini and Paul Tarau. On uniquely closable and uniquely typable skeletons of lambda terms. In *Logic-based program synthesis and transformation*, volume 10855 of *Lecture Notes in Comput. Sci.*, pages 252–268. Springer, Cham, 2018.

[14] Haskell B. Curry, Robert Feys, William Craig, J. Roger Hindley, and Jonathan P. Seldin. *Combinatory Logic*, volume 1. North-Holland Amsterdam, 1958.

[15] René David, Katarzyna Grygiel, Jakub Kozik, Christophe Raffalli, Guillaume Theyssier, and Marek Zaionc. Asymptotically almost all λ-terms are strongly normalizing. *Log. Methods Comput. Sci.*, 9(1):1:02, 30, 2013.

[16] Nicolaas G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the church-rosser theorem. *Indagationes Mathematicae (Proceedings)*, 75(5):381–392, 1972.

[17] Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer. Boltzmann samplers for the random generation of combinatorial structures. *Combin. Probab. Comput.*, 13(4-5):577–625, 2004.

[18] Philippe Flajolet, Éric Fusy, and Carine Pivoteau. Boltzmann sampling of unlabelled structures. In *Proceedings of the Ninth Workshop on Algorithm Engineering and Experiments and the Fourth Workshop on Analytic Algorithmics and Combinatorics*, pages 201–211. SIAM, Philadelphia, PA, 2007.

[19] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[20] Philippe Flajolet and Robert Sedgewick. *Analytic Ccombinatorics*. Cambridge University Press, Cambridge, 2009.

[21] Bernhard Gittenberger and Isabella Larcher. On the number of variables in special classes of random lambda-terms. In *29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms*, volume 110 of *LIPIcs. Leibniz Int. Proc. Inform.*, pages Art. No. 25, 14. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018.

[22] Clemens Grabmayer. Linear depth increase of lambda terms in leftmost-outermost rewrite sequences. In *A tribute to Albert Visser*, volume 30 of *Tributes*, pages 125–139. Coll. Publ., [London], 2016.

[23] Katarzyna Grygiel, Paweł M. Idziak, and Marek Zaionc. How big is BCI fragment of BCK logic. *J. Logic Comput.*, 23(3):673–691, 2013.

[24] Katarzyna Grygiel and Pierre Lescanne. Counting and generating terms in the binary lambda calculus. *J. Funct. Programming*, 25:e24, 25, 2015.

[25] Ryu Hasegawa. The generating functions of lambda terms (extended abstract). In *Combinatorics, complexity, & logic (Auckland, 1996)*, Springer Ser. Discrete Math. Theor. Comput. Sci., pages 253–263. Springer, Singapore, 1997.

[26] Hsien-Kuei Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19(3):329–343, 1998.

[27] Pierre Lescanne. On counting untyped lambda terms. *Theoret. Comput. Sci.*, 474:80–97, 2013.

[28] Małgorzata Moczurad, Jerzy Tyszkiewicz, and Marek Zaionc. Statistical properties of simple types. *Math. Structures Comput. Sci.*, 10(5):575–594, 2000.

[29] Zbigniew Paiłka. Testing an optimising compiler by generating random lambda terms. In *Proceedings of the 6th International Workshop on Automation of Software Test*, pages 91–97. ACM, 2011.

[30] Ryoma Sin'ya, Kazuyuki Asada, Naoki Kobayashi, and Takeshi Tsukada. Almost every simply typed λ-term has a long β-reduction sequence. In *Foundations of software science and computation structures*, volume 10203 of *Lecture Notes in Comput. Sci.*, pages 53–68. Springer, Berlin, 2017.

[31] Paul Tarau. A hiking trip through the orders of magnitude: deriving efficient generators for closed simply-typed lambda terms and normal forms. In *Logic-based program synthesis and transformation*, volume 10184 of *Lecture Notes in Comput. Sci.*, pages 240–255. Springer, Cham, 2017.

[32] Jue Wang. Generating random lambda calculus terms. 2005. Unpublished manuscript.

[33] Xuejun Yang, Yang Chen, Eric Eide, and John Regehr. Finding and understanding bugs in c compilers. In *ACM SIGPLAN Notices*, volume 46, pages 283–294. ACM, 2011.

[34] Noam Zeilberger. Linear lambda terms as invariants of rooted trivalent maps. *J. Funct. Programming*, 26:e21, 20, 2016.

[35] Noam Zeilberger and Alain Giorgetti. A correspondence between rooted planar maps and normal planar lambda terms. *Log. Methods Comput. Sci.*, 11(3):3:22, 39, 2015.

Department of Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria.
E-mail address: gittenberger@dmg.tuwien.ac.at

Department of Discrete Mathematics and Geometry, Technische Universität Wien, Wiedner Hauptstrasse 8-10/104, A-1040 Wien, Austria.
E-mail address: isabella.larcher@tuwien.ac.at