# THE NUMBER OF HUFFMAN CODES, COMPACT TREES, AND SUMS OF UNIT FRACTIONS

CHRISTIAN ELSHOLTZ, CLEMENS HEUBERGER, AND HELMUT PRODINGER

Abstract. The number of "nonequivalent" Huffman codes of length $r$ over an alphabet of size $t$ has been studied frequently. Equivalently, the number of "nonequivalent" complete $t$-ary trees has been examined. We first survey the literature, unifying several independent approaches to the problem. Then, improving on earlier work we prove a very precise asymptotic result on the counting function, consisting of two main terms and an error term.



## 1. Introduction

### 1.1. A problem in coding theory.

Let a source $S$ emit $r$ words $w_1, \ldots, w_r$ with probabilities $p_1, \ldots, p_r$ respectively. Here $0 \leq p_i \leq 1$ and $\sum_{i=1}^r p_i = 1$. For each word $w_i$ we assign a code word $c_i = c_i(w_i)$ over an alphabet of size $t$. Let $l_i$ denote the length of the codeword $c_i$. For a given source $S$, a compact code minimises the average length $\bar{l} = \sum_{i=1}^r p_i l_i$. Huffman [16] showed how to construct a code with minimum average word length, given the word probabilities $p_i$. These Huffman codes are prefix-free, and can therefore be decoded instantaneously. Moreover these codes can be found efficiently.

The Kraft-McMillan inequality states: For an alphabet of size $t$ and a source that emits $r$ words, a necessary and sufficient condition for the existence of an instantaneous code with code word lengths $l_1, \ldots, l_r$ is that

$$\sum_{i=1}^{r} \frac{1}{t^{l_i}} \leq 1. \tag{1.1}$$

Moreover, for the existence of a uniquely decipherable code inequality (1.1) is necessary.

Let us call a code *compact* if it satisfies the Kraft equality:

$$\sum_{i=1}^{r} \frac{1}{t^{l_i}} = 1. \tag{1.2}$$

When multiplying the equation by $t^{l_r}$ we observe that in a compact code the number of codewords of maximal length $l_r$ is divisible by $t$. Also, if there are two distinct codewords starting with the same prefix $a_1 \ldots a_q$ but then continuing differently, $a_1 \ldots a_q b_1 \ldots$ and $a_1 \ldots a_q b_2 \ldots$, then all $t$ possible symbols must occur at position $q+1$. In other words, if a sequence branches, it branches into all $t$ possible directions. This is the reason why it is possible to model the situation by means of a rooted $t$-ary tree, which we do below. As it is possible to arrive from a given Huffman code at a solution of equation (1.2), and vice versa, to arrive from a solution to this equation at an admissible Huffman code it is natural to consider all Huffman codes with the same set of word lengths as "equivalent"codes.

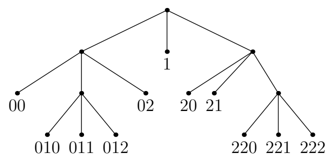
2010 *Mathematics Subject Classification.* Primary 05A16; Secondary: 05A15, 05C05, 05C30, 11D68, 68P30, 68R10, 94A10.

Example: Let $t = 3$. Let the code consist of the codewords:

$$00, 010, 011, 012, 02, 1, 20, 21, 220, 221, 222.$$

The code can be nicely represented by the tree in Figure 1.

Figure 1. Rooted tree corresponding to the code $\{00, 010, 011, 012, 02, 1, 20, 21, 220, 221, 222\}$.




Below we list a number of alternative ways of defining our main object. This reflects that the same type of question has been studied from various points of view, often without being aware of the corresponding results expressed in a different mathematical language.

We use Kraft's equality as the basis for our first definition. It stresses the number theoretic properties and was at the origin of the Boyd's [5] work.

**Definition 1** (Number theoretic definition). *Let $f_t(r)$ denote the number of solutions of the equation*

$$\sum_{i=1}^{r} \frac{1}{t^{x_i}} = 1,$$

*where the $x_i$ are nonnegative integers and $0 \leq x_1 \leq \cdots \leq x_r$.*

For more information on other counting functions related to representations of one as a sum of unit fractions, see [6] and [8].

Collecting the number of words of the same length (corresponding to $x_i$ in the last definition), one arrives at an alternative definition: From our point of view, all codes with the same number of words of a given length are equivalent. This suggests the following definition:

**Definition 2** (Huffman sequences). *Let $t \geq 2$ and $r \geq 1$ be positive integers. Let $f_t(r)$ denote the number of sequences of non-negative integers*

$$(a_0, a_1, \ldots, a_l), \quad l \geq 0,\, a_l > 0, \quad \sum_{i=0}^{l} a_i = r, \quad \sum_{i=0}^{l} \frac{a_i}{t^i} = 1.$$

1.2. **Rooted trees.** Let us recall some vocabulary from graph theory: A rooted tree is a connected cycle free graph, with one vertex being distinguished (root). (We will draw it on the top, all other vertices below). We say the tree is $t$-ary, if all those vertices, which are not the root, are either a leaf, that is an end of a path from the root, or have one predecessor and $t$ children. All non-leaves are called inner vertices. Note that the root is also an inner vertex unless for the trivial tree of order one. In other words, for the trees we consider, the root has degree $t$, all other vertices either have degree 1 (leaf) or have degree $t + 1$.

**Definition 3** (Canonical rooted tree). *A rooted tree is called canonical if its corresponding prefix code has the property that the lexicographic ordering of its words corresponds to a nondecreasing ordering of the word lengths.*

*Let us say that two rooted $t$-ary trees are equivalent, if their number of leaves at distance $i$ from the root is the same, for all $i$. Let $f_t(r)$ denote the number of equivalence classes of $t$-ary rooted trees with exactly $r$ leaves.*

Note that each equivalence class contains exactly one canonical tree. Also, if the tree has $a_i$ leaves at distance $i$ from the root, then $\sum_i \frac{a_i}{t^i} = 1$. This follows inductively, since a leaf at distance $i$ from the root, i.e. which contributes a weight $\frac{1}{t^i}$, can be split into $t$ children at distance $i+1$, of weight $\frac{1}{t^{i+1}}$ each. As these rooted $t$-ary trees correspond to a compact code, we also call these trees "compact trees".

Using Definition 3 one would for example replace the code

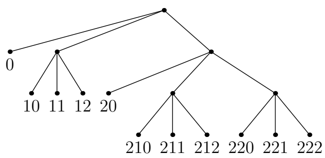
$$\{00, 010, 011, 012, 02, 1, 20, 21, 220, 221, 222\}$$

by the following equivalent code:

$$\{0, 10, 11, 12, 20, 210, 211, 212, 220, 221, 222\}.$$

The corresponding canonical rooted tree is in Figure 2. In our usual way of drawing these diagrams, a canonical tree therefore has the longer paths as far to the right hand side as possible.

FIGURE 2. Canonical tree, corresponding to $\{0, 10, 11, 12, 20, 210, 211, 212, 220, 221, 222\}$.




1.3. **A problem on bounded degree sequences.** The number $a_i$ of code words of length $i$, or leaves at level $i$ is of course bounded above by $t^i$. But there is no absolute bound on $\frac{a_i}{a_{i-1}}$. Let us study another sequence instead, namely $b_1 = 1, b_i = tb_{i-1} - a_{i-1}$, see Komlos, W. Moser and Nemetz [20] and Flajolet and Prodinger [11]. The problems of counting these sequences are equivalent to the earlier counting problem. For these sequences the ratios $\frac{b_i}{b_{i-1}}$ are bounded, which is why one may call these sequences "bounded degree sequences". Flajolet and Prodinger [11] used this definition when they counted level number sequences of trees.

**Definition 4** (Bounded degree)**.** *Let $t \geq 2, r \geq 2$ be integers. Let $f_t(r)$ denote the number of sequences*

$$(b_1, \ldots, b_l), \quad l \geq 1, \quad b_1 = 1, \quad 1 \leq b_i \leq tb_{i-1} \quad (i = 2, \ldots, l), \quad \sum_{i=1}^{l} b_i = \frac{r-1}{t-1}.$$

*For convenience we will later also use $g_t(n) = f_t(1+n(t-1))$. (Here, one can think of $n = \frac{r-1}{t-1}$).*

A bijection between the last two definitions is as follows: Given a canonical tree, we set $b_i$ to be the number of inner vertices at height $i-1$. Observe that the $b_i$ inner vertices guarantee that there are at most $tb_i$ vertices of any type (inner vertices or leaves) on the next level.

A very similar definition is due to Even and Lempel [9].

**Definition 5** (Proper words)**.** *Let $t \geq 2$ and $n \geq 1$ be integers. A word $u_1 \dots u_n$ over the alphabet $\{0,1\}$ is said to be a proper word, if it can be written in the form $u_1 \dots u_n = 0^{c_0}10^{c_1}1 \dots 0^{c_{l-1}}10^{c_l}$ such that $c_0 = 0$ and $0 \leq c_{i+1} \leq tc_i + t - 1$ holds for all $0 \leq i \leq l-1$.*

Note that the sequence $c_i$ describes the lengths of the runs of consecutive zeros. We note also that from the representation as a word of length $n$, we immediately get $\sum_{i=1}^{l} c_i = n - l$.

To see that Definition 5 is equivalent to Definition 4, we simply note that the relations $b_{i+1} = c_i + 1$ and $n = \frac{r-1}{t-1}$ induce a bijection between the objects counted in the two definitions. Even and Lempel [9] also give a combinatorial interpretation of this bijection (for $t = 2$, but the generalisation is straight-forward): essentially, for each 1 in a proper word, they replace a leaf of maximum height by an inner vertex with $t$ leaves as successors; for each 0, they replace a leaf of second-most height by an inner vertex with $t$ leaves as successors.

We briefly mention some further approaches which investigate equivalent sequences. Working on a different problem, Minc [22] reduced it to the study of a binary bounded degree sequence, Definition 4 above. Let $A$ be a free commutative entropic cyclic groupoid. The number of elements of $A$ of a given degree turns out to satisfy the relation above. (For a full description we must refer to [22]). The condition in Definition 4 looks like a special partition function. Andrews [2] expanded on Minc's work, in particular studying generating functions.

A further problem, on lambda algebras $\Lambda_p$, has been related to these sequences, see Tangora [31].

1.4. **An example.** As an example for these various definitions, let us compute $f_2(5) = 3$ in the different forms. Using Definition 1:

$$1 = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{2} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8}$$

is a complete list of all solutions.

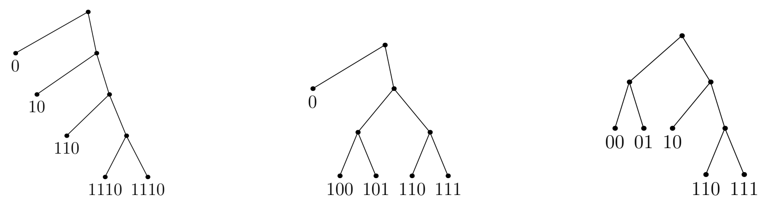
Counting Huffman sequences (Definition 2) we count $(a_0, a_1, \dots)$ where $a_i$ is the number of occurrences of the fraction $\frac{1}{t^i}, i \geq 1$. Here with $t = 2$ these sequences are:

$$(0, 1, 1, 1, 2), (0, 1, 0, 4), (0, 0, 3, 2).$$

Let us explicitly write down the compact Huffman codes.

$$C_1 = \{0, 10, 110, 1110, 1111\},\ C_2 = \{0, 100, 101, 110, 111\},\ C_3 = \{00, 01, 10, 110, 111\}.$$

The bounded degree sequences counted in Definition 4 are $(1,1,1,1)$, $(1,1,2)$, $(1,2,1)$. The proper words in Definition 5 are $(111), (110), (101)$. The canonical trees (Definition 3) are the following:



1.5. **An observation.** When evaluating $f_t(r)$, according to the Definition 2 of Huffman sequences it suffices to investigate in which way a solution counted by $f_t(r-t+1)$ can be split. Let $S_t(r)$ denote the set of all sequences counted by $f_t(r)$. Generally, $(a_0, a_1, \dots, a_i, \dots, a_l)$ can be split into $(a_0, a_1, \dots, a_i - 1, a_i + t, \dots, a_l)$, whenever $a_i > 0$. Starting from a complete set of solutions, that is $S_t(r-t+1)$, one only needs to branch each sequence at the last two positions, in order to compile a complete set of solutions, $S_t(r)$. The reason for this is that all elements of $S_t(r)$ obtained from branching at any of the earlier positions will be obtained from another member of $S_t(r-t+1)$ by branching at the last two positions. Before we generally prove this let us look at an example. Let us determine $S_2(6)$, starting from the three elements of $S_2(5) = \{01112, 0104, 0032\}$:

$$0111|2 \to 0111|12,\ 011|12 \to 011|04,\ 010|4 \to 010|32,\ 003|2 \to 003|12,\ 00|32 \to 00|24.$$

There is no need to consider

$$0|1112 \to 0|0312, \text{ or } 01|112 \to 01|032 \text{ or } 0|104 \to 0|024,$$

as these are obtained otherwise.

To see this generally, let us consider the step from $f_t(r-t+1)$ to $f_t(r)$: If $(a_0, a_1, a_2, a_3, \dots, a_l) \in S_t(r-t+1)$, i.e. $\sum_{i=0}^{l} a_i = r - t + 1$, with $a_l > 0$, we need to check if $(a_0, a_1, \dots, a_i - 1, a_{i+1} + t, a_{i+2}, \dots, a_l) \in S_t(r)$ will be reached by branching an appropriate element of $S_t(r-t+1)$ in any of the last two positions only.

Note that $(a_0, a_1, \dots, a_i - 1, a_{i+1} + t, a_{i+2}, \dots, a_{l-1} + 1, a_l - t) \in S_t(r-t+1)$. Hence one reaches $(a_0, a_1, \dots, a_i - 1, a_{i+1} + t, a_{i+2}, \dots, a_{l-1}, a_l) \in S_t(r)$ by branching in the last two positions only. We may also observe that this gives a trivial upper bound of $f_t(r) \le 2^{\frac{r-1}{t-1}}$.

Using the above observation of branching at two positions only, Narimani and Khosravifard [23] describe a recursive algorithm to create all codes counted by $f_t(r)$.

The first terms of the sequence $f_2(r)$ are:

$$t = 2 : 1, 1, 1, 2, 3, 5, 9, 16, 28, 50, 89, 159, 285, 510, 914, 1639, \dots$$

The values of $f_3(r)$ are zero, whenever $r$ is even. The nontrivial part of the sequence for odd $r$, that is $g_3(n)$ starts with

$$t = 3 : 1, 1, 1, 2, 4, 7, 13, 25, 48, 92, 176, \dots$$

(see also [28]). For general $t$, the sequence is only non-zero for $r = 1 + (t-1)n$. For convenience one examines $g_t(n) = f_t(1 + n(t-1))$ instead, see Definition 4. For reference purposes we list the first values of the sequences $g_t(n)$ in Table 1. In these tables one can easily notice the observation above, $g_t(n) = f_t(r) \le 2^{\frac{r-1}{t-1}} = 2^n$.

The sequences $g_2(n), g_3(n)$ and $g_4(n)$ have been included into the OEIS (sequences A002572, A176485 and A176503). (The latter two sequences only after the appearance of the Paschke et al. paper [28].)

1.6. **The growth of $f_t(r)$.** As far as we are aware of, Bende (1967) [4] and Norwood (1967) [24] were the first to examine the sequence $f_2(r)$, and they observed the connection to coding theory and trees. (Minc's 1958 paper [22] was, of course, earlier but had less interest in the sequence itself.) Bende asked about the asymptotic growth. Erdős in his review of Bende's paper (Mathematical Reviews) also wrote it is "desirable" to know the asymptotic.

| $t$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 1 | 1 | 2 | 3 | 5 | 9 | 16 | 28 | 50 | 89 | 159 | 285 | 510 | 914 | 1639 | 2938 | 5269 | 9451 | 16952 |
| 3 | 1 | 1 | 1 | 2 | 4 | 7 | 13 | 25 | 48 | 92 | 176 | 338 | 649 | 1246 | 2392 | 4594 | 8823 | 16945 | 32545 | 62509 |
| 4 | 1 | 1 | 1 | 2 | 4 | 8 | 15 | 29 | 57 | 112 | 220 | 432 | 848 | 1666 | 3273 | 6430 | 12632 | 24816 | 48754 | 95783 |
| 5 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 31 | 61 | 121 | 240 | 476 | 944 | 1872 | 3712 | 7362 | 14601 | 28958 | 57432 | 113904 |
| 6 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 32 | 63 | 125 | 249 | 496 | 988 | 1968 | 3920 | 7808 | 15552 | 30978 | 61705 | 122910 |
| 7 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 127 | 253 | 505 | 1008 | 2012 | 4016 | 8016 | 16000 | 31936 | 63744 | 127234 |
| 8 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 255 | 509 | 1017 | 2032 | 4060 | 8112 | 16208 | 32384 | 64704 | 129280 |
| 9 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 511 | 1021 | 2041 | 4080 | 8156 | 16304 | 32592 | 65152 | 130240 |
| 10 | 1 | 1 | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1023 | 2045 | 4089 | 8176 | 16348 | 32688 | 65360 | 130688 |

Table 1. Values of $g_t(n)$ for $2 \le t \le 10$ and $1 \le n \le 20$.

The early 1970's saw a considerable number of contributions to the problem, such as Boyd [5], Even and Lempel [9], and Gilbert [13].

A trivial upper bound for the number of rooted canonical trees on $|V|$ vertices is $2^{\binom{|V|}{2}}$. A much more precise bound is the number of all trees. The number of binary trees on $|V|$ vertices is determined by the Catalan numbers $\frac{1}{n+1}\binom{2n}{n} = O(4^n n^{-3/2})$ and the number of non-isomorphic trees is asymptotically $\sim C_2\, C_1^n n^{-5/2}$, where $C_1 = 2.955\ldots$ and $C_2 = 0.5349\ldots$, see Otter [26].

A trivial lower bound comes from observing that Definition 4 shows that $f_2(r) \ge F_r$, where $F_r$ is the number of ways of partitioning $r-1$ into ones and twos. It is known that this is the $r$-th Fibonacci number so that $f_2(r) \ge 0.4472 \times 1.618329^r$ (for sufficiently large $r$). Similarly, a lower bound on $f_t(r)$, can be obtained by partitioning $r-1$ into 1's, 2's ... and $t$'s. By means of the generating series of $\frac{1}{1-z-z^2-\cdots-z^t}$ and determining a real root of the equation $1-z-z^2-\cdots-z^t=0$ near 0.5 the corresponding generalised Fibonacci number $F_{t,r}$ can be shown to be about $c_t\rho_t^r$, where $\rho_t \approx 2-\frac{1}{2^t-\frac{t}{2}}$, and $c_t$ is a positive constant. In the next section we will refine an analysis of this type considerably.

Boyd (1975) [5], Komlos, W. Moser and Nemetz (1984) [20], Flajolet and Prodinger (1987) [11], all independently, gave an asymptotic:

$$f_2(r) \sim R\rho^r,$$

where $R \approx 0.14185$, $\rho \approx 1.7941471$. Boyd and Flajolet and Prodinger additionally gave an error term: $f_2(r) = R\rho^r + O(\tilde{\rho}^r)$, where Boyd proves $\tilde{\rho} = 1.55$, and Flajolet and Prodinger proved that this even holds for $\tilde{\rho} = \frac{10}{7}$. Boyd, and Komlos, W. Moser and Nemetz also study the case of more general $t$. As noted before: as $f_t(r)$ is positive only for $r = 1 + n(t-1)$, one examines $g_t(n) = f_t(1+n(t-1))$ instead.

In particular Komlos, Moser and Nemetz observed that $g_t(n) \sim K_t\rho_t^n$ with $\rho_t \to 2$, as $t$ increases. Flajolet and Prodinger [11] also refer to other areas, where the sequence $f_2(r)$ naturally occurs.

Building upon [11], but not being aware of [5] nor [20], Tangora (1991) [31] generalised the results to prime values of $t$.

Another string of references follows from Gilbert's experimental observation that $f_2(r) \approx 0.148(1.791)^r$, see [13]. The observation was based on the values for $r \le 30$, and is relatively close to the true asymptotic $f_2(r) \sim 0.1418\ldots(1.7941\ldots)^r$. However, these approximations have been referred to in the more recent coding literature, see for example [27], [29], [1], [23], [18] and [19].

More recently Burkert (2010) [7] and Paschke, Burkert, Fehribach (2011) [28] studied $f_2(r)$ and $f_t(r)$ respectively, unfortunately with inferior results and unfortunately being unaware of the earlier work.[1]

In the results that we describe in detail in the next section, we state a rather precise asymptotic formula, with two main terms, and an error term, which is *exponentially* smaller. As an example, one finds an approximation

$$f_2(n+1) \approx R\rho^{n+1} + R_2\rho_2^{n+1},$$

with

$$\begin{aligned} \rho &= 1.794147187541686, & \rho_2 &= 1.279549134726681, \\ R &= 0.1418532020854094, & R_2 &= 0.0612410410312. \end{aligned}$$

Let us evaluate $f_2(50) \approx 699427308155.394\ldots$. While the error analysis of Theorem 7 (below) gives an error of $|f_2(50) - (R\rho^{50} + R_2\rho_2^{50})| \leq 36.6 \cdot 1.123^{50} \leq 12092$, the absolute error is much smaller and, in this case, the above approximation predicts the *correct* value of $f_2(50) = 699427308155$.

1.7. **A note on algorithms and complexity.** The question of the complexity of the evaluation of $f_2(r)$ is raised in Even and Lempel [9]. They give an algorithm to determine $f_2(r)$ in $O(r^3)$ additions. This appears to be the only algorithm with analysis of its complexity. They also state another algorithm to give a complete list of the $f_2(r)$ elements.

Huffman, Johnson and Wilson [15] describe another algorithm to give a complete list.

A tree based algorithm for generating the binary compact codes is described in [18]. Narimani and Khosravifard [23] describe a recursive algorithm to create all $t$-ary codes of length $r$ by those of length $r - t + 1$.

## 2. Results

In the following, a tree will always be a $t$-ary rooted canonical tree. The set of $t$-ary canonical trees is denoted by $\mathcal{T}$. The number of inner vertices (non-leaves) of a tree $T$ is denoted by $n(T)$. Setting $c_n := g_t(n)$ to be the number of trees $T \in \mathcal{T}$ with $n$ inner vertices, we are interested in the generating function

$$F(q) = \sum_{n\geq 0} c_n q^n = \sum_{T\in\mathcal{T}} q^{n(T)}.$$

This generating function can be computed explicitly:

[1]The oversights some decades ago can be easily explained due to the fact that the results were discovered independently by people with interests in number theory, coding theory or graph theory. Boyd's paper [5] has a number theoretic title, the Komlos et al. paper [20] a coding title and appeared in a less accessible journal. Using standard tools such as MathSciNet, Zentralblatt, Google Scholar, Online Encyclopedia of Integer Sequences (OEIS) we found a considerable corpus of literature referring to the result that $f_t(r) \sim K_t \cdot \rho_t^r$.

**Theorem 6.** *Setting* $[k] := 1 + t + t^2 + \cdots + t^{k-1}$, *we have*

$$F(q) = \frac{\displaystyle\sum_{j=0}^{\infty}(-1)^j q^{[j]} \prod_{i=1}^{j} \frac{q^{[i]}}{1-q^{[i]}}}{\displaystyle\sum_{j=0}^{\infty}(-1)^j \prod_{i=1}^{j} \frac{q^{[i]}}{1-q^{[i]}}}.$$

Using the generating function, we can give a very precise asymptotic expression for $c_n$. In view of the numerous asymptotic approximations we would like to point out that this is the first result containing two main terms and an explicit error term.

**Theorem 7.** *For* $t \geq 2$, *the following holds:*

$$c_n = g_t(n) = R\rho^{n+1} + R_2\rho_2^{n+1} + R_3 r_3^n \varepsilon_1(t,n), \tag{2.1}$$

*Here* $\rho > \rho_2 > r_3$ *and* $R, R_2, R_3$ *are positive real constants to be specified below, and depending on* $t$. *Here and below,* $\varepsilon_j(\ldots)$, $j = 1, \ldots,$ *denote real functions with* $|\varepsilon_j(\ldots)| \leq 1$ *for all valid values of the respectively indicated parameters.*

*For* $t \geq 16$ *we have*

$$\rho = 2 - \frac{1}{2^{t+1}} - \frac{t+3}{2^{2t+3}} - \frac{3t^2+19t+24}{2^{3t+6}} + \frac{0.28t^3}{2^{4t}}\varepsilon_2(t), \tag{2.2}$$

$$\begin{split}\rho_2 = 1 + \frac{\log 2}{t} - \frac{\log 2 - \log^2 2}{2t^2} + \frac{4\log^3 2 + 3\log^2 2 + 6\log 2}{24t^3} \\ + \frac{2\log^4 2 + 54\log^3 2 - 27\log^2 2 - 6\log 2}{48t^4} + \frac{0.26}{t^5}\varepsilon_3(t),\end{split} \tag{2.3}$$

$$r_3 = 1 + \frac{\log 2}{t} - \frac{\log 2 - \log^2 2}{2t^2}, \tag{2.4}$$

$$R = \frac{1}{8} + \frac{t-2}{2^{t+5}} + \frac{2t^2+3t-5}{2^{2t+7}} + \frac{9t^3+45t^2+20t-68}{2^{3t+10}} + \frac{t^4}{50\cdot 2^{4t}}\varepsilon_4(t), \tag{2.5}$$

$$R_2 = \frac{1}{4t} - \frac{4\log 2 + 1}{8t^2} + \frac{0.77}{t^3}\varepsilon_5(t), \tag{2.6}$$

$$R_3 = 5t^4. \tag{2.7}$$

*For* $3 \leq t \leq 15$, (2.1) *holds with* (2.2), (2.5), (2.6) *and the values for* $\rho_2, r_3$ *and* $R_3$ *given in Table 2.*

*For* $t = 2$, (2.1) *holds with* (2.6) *and the values for* $\rho$, $\rho_2$, $r_3, R$ *and* $R_3$ *given in Table 2.*

For simplicity the functions $\varepsilon_j$ can be thought of as $O(1)$ terms. Some of our proofs indeed depend on explicit values of the error bounds. For this reason we had to compute absolute $O$-constants in any case, and decided to include these in the statement of the theorem.

The asymptotic result focusses on the first and the second exponential terms $\rho^{n+1}$ and $\rho_2^{n+1}$ and no effort has been made to improve the error term $r_3^n$: note that for large $t$ it is not much smaller then the second order term $\rho_2^{n+1}$. For Table 2 the values $r_3$ have been improved by a computer calculation in comparison with Equation (2.4), also leading to a stronger value of the constant $R_3$ in comparison with (2.7). In principle, this type of improvement is possible for any fixed $t \geq 16$ as well.

| $t$ | $\rho$ | $\rho_2$ | $r_3$ | $R$ | $R_2$ | $R_3$ |
|---|---|---|---|---|---|---|
| 2 | 1.794147187541686 | 1.279549134726681 | 1.123 | 0.1418532020854094 | 0.0612410410312$*$ | 36.6 |
| 3 | 1.920712538405631$*$ | 1.211479378117327 | 1.098 | 0.1338681353605138$*$ | 0.05040725710011751$*$ | 39.0 |
| 4 | 1.964624757813775$*$ | 1.165158374565692 | 1.083 | 0.1305243270109503$*$ | 0.04239969309700251$*$ | 58.4 |
| 5 | 1.983293986764127$*$ | 1.134459698442781 | 1.074 | 0.1284678647212778$*$ | 0.03633182386516354$*$ | 70.7 |
| 6 | 1.991897175722647$*$ | 1.113019849812048 | 1.068 | 0.1271299952558400$*$ | 0.03168855397536632$*$ | 50.0 |
| 7 | 1.996015107731262$*$ | 1.097324075593615 | 1.063 | 0.1262776860399922$*$ | 0.02807600275247040$*$ | 59.6 |
| 8 | 1.998025544625657$*$ | 1.085389242111509 | 1.059 | 0.1257503987658994$*$ | 0.02520568904841775$*$ | 48.1 |
| 9 | 1.999017663916874$*$ | 1.076032488551186 | 1.056 | 0.1254328058843682$*$ | 0.02287594728315024$*$ | 24.0 |
| 10 | 1.999510161506312$*$ | 1.068511410911158 | 1.053 | 0.1252458295005635$*$ | 0.02094759256441895$*$ | 19.7 |
| 11 | 1.999755441055006$*$ | 1.062339511503337$*$ | 1.050 | 0.1251378340222618$*$ | 0.01932397366876184$*$ | 20.1 |
| 12 | 1.999877817773010$*$ | 1.057186165846774$*$ | 1.047 | 0.1250764428075050$*$ | 0.01793689446751572$*$ | 26.6 |
| 13 | 1.999938935019296$*$ | 1.052819586914068$*$ | 1.044 | 0.1250420050254539$*$ | 0.01673722535920120$*$ | 80.6 |
| 14 | 1.999969474502513$*$ | 1.049072853620226$*$ | 1.042 | 0.1250229006766309$*$ | 0.01568876914448585$*$ | 43.3 |
| 15 | 1.999984739115025$*$ | 1.045822904924682$*$ | 1.040 | 0.1250124013324635$*$ | 0.01476426249364319$*$ | 39.0 |

TABLE 2. Values for small values of $t$. Starred ($*$) entries correspond to values satisfying the asymptotic estimates of Theorem 7. The values could be given with much higher precision, there is some uncertainty about the last digit.

The asymptotic expansions of $\rho$, $\rho_2$, $R$ and $R_2$ can always be refined by further iterating the fixed point equations in the proof of Proposition 10. So for fixed $k$, we could refine the estimates for $\rho$ and $R$ to a precision of $t^k 2^{-tk}$ and the estimates for $\rho_2$ and $R_2$ to a precision of $t^{-k}$.

## 3. Generating Function

This section is devoted to the proof of Theorem 6.

*Proof of Theorem 6.* In the proof of the theorem, we will actually consider more refined statistics in order to derive a functional equation for a more general generating function.

The height of a vertex in a rooted tree is defined to be its distance from the root. So the root has height 0. The height $\mathsf{height}(T)$ of a tree $T$ is defined to be the maximal height of its vertices.

For a rooted tree $T$, we set $m(T)$ to be the number of leaves of maximum height of $T$.

We will derive a functional equation for the generating function

$$G(q,u) = \sum_{T\in\mathcal{T}} q^{n(T)} u^{m(T)},$$

i.e., $u$ counts the number of leaves of maximal height and $q$ counts the number of inner vertices. By definition, we have $F(q) = G(q,1)$.

To derive the functional equation for $G(q,u)$, we partition $\mathcal{T}$ with respect to the height and consider

$$G_k(q,u) = \sum_{\substack{T\in\mathcal{T}\\ \mathsf{height}(T)=k}} q^{n(T)} u^{m(T)}.$$

Obviously, we have

$$G(q,u) = \sum_{k\geq 0} G_k(q,u).$$

A tree $T$ of height $k$ corresponds to exactly $m(T)$ trees $T'_j$, $j \in \{1, \dots, m(T)\}$, of height $k+1$: $T'_j$ arises from $T$ by replacing $j$ of the $m(T)$ leaves of maximum height by vertices with $t$ attached leaves. On the other hand, all trees $T'$ of height $k+1$ are uniquely described by this process.

Thus we have

$$\begin{aligned} G_{k+1}(q,u) &= \sum_{\substack{T \in \mathcal{T} \\ \mathsf{height}(T)=k}} \sum_{j=1}^{m(T)} q^{n(T)+j} u^{jt} \\ &= \sum_{\substack{T \in \mathcal{T} \\ \mathsf{height}(T)=k}} q^{n(T)} \cdot qu^t \cdot \frac{1-(qu^t)^{m(T)}}{1-qu^t} \\ &= \frac{qu^t}{1-qu^t} \left( G_k(q,1) - G_k(q,qu^t) \right). \end{aligned} \tag{3.1}$$

We have $G_0(q,u) = u$, so summing over all $k \geq 0$ yields

$$G(q,u) - u = \frac{qu^t}{1-qu^t} (G(q,1) - G(q,qu^t)). \tag{3.2}$$

The generating function $G(q,u)$ is certainly convergent for $|u| \leq 1$ and $|q| < 1/2$, as can be seen from (3.1).

We now keep $q$ with $|q| < 1/2$ fixed and consider everything as a function of $u$ with $|u| \leq 1$. We use the abbreviations $h(u) = qu^t/(1-qu^t)$ and $g(u) = G(q,u)$. We rewrite the functional equation (3.2) as

$$g(u) = u + h(u)g(1) - h(u)g(qu^t).$$

By iteration, we obtain

$$\begin{aligned} g(u) &= a_k(u) + b_k(u)g(1) + c_k(u)g(q^{[k+1]}u^{t^{k+1}}), \\ a_k(u) &= \sum_{j=0}^{k} (-1)^j q^{[j]} u^{t^j} \prod_{i=0}^{j-1} h(q^{[i]}u^{t^i}), \\ b_k(u) &= \sum_{j=0}^{k} (-1)^j \prod_{i=0}^{j} h(q^{[i]}u^{t^i}), \\ c_k(u) &= (-1)^{k+1} \prod_{i=0}^{k} h(q^{[i]}u^{t^i}) \end{aligned}$$

for $k \geq 0$. As $|h(u)| \leq \frac{|q|}{1-|q|} < 1$ holds for all $|u| \leq 1$, the limits

$$\begin{aligned} a(u) &= \sum_{j=0}^{\infty} (-1)^j q^{[j]} u^{t^j} \prod_{i=0}^{j-1} h(q^{[i]}u^{t^i}), \\ b(u) &= \sum_{j=0}^{\infty} (-1)^j \prod_{i=0}^{j} h(q^{[i]}u^{t^i}) \end{aligned}$$

exist and we have $\lim_{k\to\infty} c_k(u) g(q^{k+1}u^{t^{k+1}}) = 0$.

Thus we obtained

$$g(u) = a(u) + b(u)g(1).$$

Setting $u = 1$ yields

$$F(q) = G(q, 1) = g(1) = \frac{a(1)}{1 - b(1)}.$$

□

## 4. Asymptotics

We will use the following notations in order to work with the generating function $F$:

$$f_j(q) = \frac{q^{[j]}}{1 - q^{[j]}},$$

$$N_K(q) = \sum_{0 \le k < K} (-1)^k q^{[k]} \prod_{j=1}^{k} f_j(q), \qquad D_K(q) = \sum_{0 \le k < K} (-1)^k \prod_{j=1}^{k} f_j(q),$$

$$N(q) = \sum_{0 \le k} (-1)^k q^{[k]} \prod_{j=1}^{k} f_j(q), \qquad D(q) = \sum_{0 \le k} (-1)^k \prod_{j=1}^{k} f_j(q).$$

The quantities have been defined such that $F(q) = N(q)/D(q)$.

We intend to work with the finite sums $D_K$ and $N_K$ for fixed values of $K$, so we need upper bounds for the approximation errors.

**Lemma 8.** *Let $K \ge 0$ and $|q|^{[K+1]} < 1/2$. Then*

$$|N(q) - N_K(q)| \le \left( \frac{1 - |q|^{[K+1]}}{1 - 2|q|^{[K+1]}} \prod_{j=1}^{K} \frac{1}{1 - |q|^{[j]}} \right) |q|^{[K] + \sum_{j=1}^{K} [j]}, \tag{4.1a}$$

$$|D(q) - D_K(q)| \le \left( \frac{1 - |q|^{[K+1]}}{1 - 2|q|^{[K+1]}} \prod_{j=1}^{K} \frac{1}{1 - |q|^{[j]}} \right) |q|^{\sum_{j=1}^{K} [j]}. \tag{4.1b}$$

*These bounds are decreasing in $t$ and increasing in $|q|$.*

*Proof.* As $|f_j(q)| \le f_j(|q|)$ and $f_j(|q|)$ is decreasing in $j$, we have

$$\begin{aligned} |D(q) - D_K(q)| &\le \sum_{k=K}^{\infty} \prod_{j=1}^{K} f_j(|q|) \prod_{j=K+1}^{k} f_j(|q|) \\ &\le \prod_{j=1}^{K} f_j(|q|) \sum_{k=K}^{\infty} f_{K+1}(|q|)^{k-K} \\ &= \frac{1}{1 - f_{K+1}(|q|)} \prod_{j=1}^{K} f_j(|q|), \end{aligned}$$

which, upon inserting the definition of $f_j$, yields (4.1b). The approximation bound (4.1b) for the numerator follows along the same lines, we get an additional factor $q^{[K]}$. □

We will also need estimates for the derivative $D'(q)$:

**Lemma 9.** *Let $t \geq 30$ and $q \in \mathbb{C}$ with $1/2 \leq |q| \leq 1/r_3$, where $r_3$ is defined in* (2.4).

*Then*

$$|D'(q) - D_4'(q)| \leq \frac{1}{2^{t^2}}.$$

*Proof.* Let $q = 1/z$ with $r_3 \leq |z| \leq 2$. Then $f_j(q) = f_j(1/z) = \frac{1}{z^{[j]}-1}$ and $|f_j(q)| = 1/|z^{[j]} - 1| \leq 1/(r_3^{[j]} - 1)$. By estimating the relevant power series, we get

$$\begin{aligned}
r_3 - 1 &\geq \frac{1}{2t},\\
r_3^{[2]} - 1 &= \exp\left((1+t)\log\left(1 + \frac{\log 2}{t} - \frac{\log 2 - \log^2 2}{2t^2}\right)\right) - 1 \geq 1,\\
r_3^{[3]} - 1 &\geq 2^t, \qquad (4.2a)\\
r_3^{[4]} - 1 &= 2^{t^2+t/2}. \qquad (4.2b)
\end{aligned}$$

We have

$$\begin{aligned}
|D'(1/z) - D_4'(1/z)| &\leq |z| \sum_{k=4}^{\infty} \prod_{j=1}^{k} f_j(1/|z|) \left(\sum_{j=1}^{k} \frac{[j]}{1 - (1/|z|)^{[j]}}\right)\\
&\leq 2 \sum_{k=4}^{\infty} \frac{t}{2^{-1+t(k-1)/2+(k-3)t^2}} \left(4t + 4\sum_{j=2}^{k}[j]\right) \leq \sum_{k=4}^{\infty} \frac{kt^{k+1}}{2^{(k-3)t^2+t(k-1)/2-4}}\\
&\leq \frac{1}{2}\sum_{k=4}^{\infty} \frac{1}{2^{t^2(k-3)}} \leq \frac{1}{2^{t^2}}.
\end{aligned}$$

□

The exponential growth of the coefficients $c_n$ of $F(q)$ is directly related to the dominating pole $1/\rho$ of $F(q)$. So we now investigate the location of the poles of $F(q)$.

**Proposition 10.** *Let $t \geq 2$. Then there are exactly two poles $1/\rho$ and $1/\rho_2$ of $F(q)$ with $|q| \leq 1/r_3$, where $r_3$ has been defined in* (2.4) *(or Table 2 for $t \in \{2, 3\}$).*

*Both $1/\rho$ and $1/\rho_2$ are simple poles of $F(q)$. The dominant pole $1/\rho$ of $F(q)$ is asymptotically given by* (2.2) *(or Table 2 for $t = 2$).*

*The residue of $F(q)$ at $1/\rho$ is $-R$ where $R$ is asymptotically given by* (2.5) *(or Table 2 for $t = 2$).*

*The pole $1/\rho_2$ is given by* (2.3) *(or Table 2 for $2 \leq t \leq 15$), the residue of $F(q)$ at $1/\rho_2$ is $-R_2$, where $R_2$ is given in* (2.6).

*Finally, we have*

$$|F(q)| \leq 5t^4 \qquad (4.3)$$

*for all $q$ with $|q| = 1/r_3$.*

The proof of Proposition 10 relies on rewriting the equation $D(q) = 0$ into two fixed point equations, one for each of the two poles. Inserting preliminary bounds into these fixed point equations improves these bounds. This method is known as bootstrapping. The first pole is an attracting fixed point of the first fixed point formulation, whereas the second pole is a repellent fixed point of this first fixed point formulation. So we need to take inverses in order to turn the second pole into an attracting fixed point. However, inversion involves extracting a $(t+1)$-st root, so several branches occur. Additional inequalities are required in order to decide which branch to take. We repeatedly use power series estimates in order to get the required inequalities. In order to sharpen these estimates, we assume that $t \geq 30$.

*Proof.* In the proof of this proposition, some more functions $\varepsilon_j(\ldots)$ occur. We first allow complex values for the $\varepsilon_j(\ldots)$, it will later turn out that those occurring in Theorem 7 have only real values.

In the following, we consider the case $t \geq 30$. Assume that $1/z$ is a pole of $F(q)$ with $|z| \geq 1 + a/t$ for some $2 \geq a \geq \log 2$. As $N(q)$ is holomorphic for $|q| < 1$, cf. Lemma 8, $1/z$ must be a root of $D(q)$. Using $K = 3$, we get

$$0 = 1 - \frac{1}{z-1} + \frac{1}{z-1}\frac{1}{z^{t+1}-1} + (D(1/z) - D_3(1/z)),$$

which is equivalent to

$$2 - z = \frac{1}{z^{t+1}-1} + (z-1)(D(1/z) - D_3(1/z)). \tag{4.4}$$

Taking absolute values, (4.1b) yields

$$2 - |z| \leq |2-z| \leq \frac{1}{|z|^{[2]}-1}\left(1 + \frac{1}{|z|^{[3]}-1}\cdot\frac{1}{1-\frac{1}{|z|^{[4]}-1}}\right). \tag{4.5}$$

We have

$$\begin{aligned}|z|^{[2]} &\geq \left(1+\frac{a}{t}\right)^{t+1} = \exp\left((t+1)\log\left(1+\frac{a}{t}\right)\right) \geq \exp\left((t+1)\left(\frac{a}{t}-\frac{a^2}{2t^2}\right)\right)\\ &= \exp\left(a + \frac{a-a^2/2}{t} - \frac{a^2}{2t^2}\right) \geq \exp\left(a+\frac{b}{t}\right) \geq e^a\left(1+\frac{b}{t}\right)\end{aligned} \tag{4.6}$$

for $b = a - 31a^2/60 > 0$. By (4.2a) and (4.2b), we have

$$\frac{1}{|z|^{[3]}-1}\cdot\frac{1}{1-\frac{1}{|z|^{[4]}-1}} \leq \frac{1.00001}{2^t}. \tag{4.7}$$

Consider now the case $a = \log 2$. Then (4.5), (4.6) and (4.7) yield

$$2 - |z| \leq \frac{1}{1+\frac{2b}{t}}\left(1 + \frac{1.00001}{2^t}\right) \leq 1 - \frac{4}{5t}. \tag{4.8}$$

We conclude that $|z| \geq 1 + \frac{4}{5t}$. So using now $a = 4/5$, (4.5), (4.6) and (4.7) yield

$$2 - |z| \leq \frac{1}{e^{4/5}-1}\left(1+\frac{1.00001}{2^t}\right) \leq 0.82$$

and therefore $|z| \geq 1.18$. Inserting this and (4.7) in (4.5) now yields

$$2 - |z| \leq |2-z| \leq \frac{1}{1.18^{t+1}-1}\left(1+\frac{1.00001}{2^t}\right) \leq \frac{0.86}{1.18^t}.$$

We conclude that $z = 2 + O(1.18^{-t})$. We now rewrite (4.4) as

$$z = 2 - \frac{1}{z^{t+1}-1} + O(2^{-t^2}). \tag{4.9}$$

Inserting $z = 2 + O(1.18^{-t})$ in the right-hand side of (4.9) yields

$$z = 2 - \frac{1}{(2 + O(1.18^{-t}))^{t+1} - 1} = 2 - \frac{1}{2^{t+1}} \left(1 + O(t\, 1.18^{-t})\right).$$

We now repeat the process: We insert this estimate in the right-hand side of (4.9) and get a better estimate. After a few iterations (and taking care of all implicit constants), we finally get (2.2). Inserting the lower and the upper bounds of (2.2) into $D_3(q)$ (and taking into account $D(q) - D_3(q)$), we see that $D(q)$ changes sign within the interval, so there is certainly a root $1/z$ of $D(q)$ fulfilling (2.2).

Inserting this asymptotic expression into $D'(q)$ and using Lemma 9, we get

$$|D'(1/z) + 4| \le 1.04t2^{-t} \tag{4.10}$$

for $t \ge 30$. This shows that there is at most one zero of $D(1/z)$ within the bounds of the asymptotic expression (2.2): if there were two, say $1/z_1$ and $1/z_2$, then

$$\begin{aligned} 4\left|\frac{1}{z_2} - \frac{1}{z_1}\right| &= \left|D(1/z_2) - D(1/z_1) + 4\left(\frac{1}{z_2} - \frac{1}{z_1}\right)\right| \\ &= \left|\int_{[1/z_1, 1/z_2]} (D'(q) + 4)\, dq\right| \le 1.04t2^{-t}\left|\frac{1}{z_2} - \frac{1}{z_1}\right|, \end{aligned}$$

which implies $1/z_1 = 1/z_2$. Here, we integrate over the straight line from $1/z_1$ to $1/z_2$. The estimate (4.10) also shows that there can only be a simple root. Thus we have shown that the only root $1/z$ of $D$ with $|z| \ge 1 + \log 2/t$ is a simple zero with $z$ as in (2.2). The residue (2.5) follows upon inserting (2.2) into $N(1/z)/D'(1/z)$. Note that this also shows that the dominant zero of the denominator does not cancel out against a zero of the numerator.

Now assume that $|D(1/z)| \le 1/t^3$ holds for some $z$ with $r_3 \le |z| \le 1 + \log 2/t$. Inserting these bounds into (4.5), we get

$$|z - 2| \le 1 - \frac{\log 2}{t} + \frac{4\log^3 2 - 3\log^2 2 + 12\log 2}{12t^2} + \frac{1.5}{t^3}\varepsilon_6(t, z) =: r'. \tag{4.11}$$

The intersection point with positive imaginary part of the circle of radius $1 + \log 2/t$ centred at the origin with the circle of radius $r'$ centred at 2 is denoted by $\xi$. We obtain

$$\xi = 1 + \frac{4\log 2 + i\sqrt{\frac{16}{3}\log^3 2 - 4\log^2 2 + 16\log 2}}{4t} + \frac{2.23}{t^2}\varepsilon_7(t).$$

In particular, we have

$$|z - 1| \le |\xi - 1| \le \frac{1.14}{t} \tag{4.12}$$

and

$$|\arg(z)| \le |\arg \xi| \le |\log \xi| \le \frac{1.18}{t}. \tag{4.13}$$

As $|D(1/z)| \le 1/t^3$, we have (after multiplication with $z - 1$)

$$0 = z - 2 + \frac{1}{z^{t+1} - 1} + \frac{2.01}{t^3}\varepsilon_8(t, z).$$

Solving for $z^{t+1}$ yields

$$z^{t+1} = 1 + \frac{1}{2 - z - \frac{2.01}{t^3}\varepsilon_8(t,z)}.$$

As $z = 1 + \frac{1.14}{t}\varepsilon_9(t,z)$ by (4.12), we obtain

$$z^{t+1} = 2 + \frac{1.19}{t}\varepsilon_{10}(t,z).$$

We conclude that

$$z = \exp\left(\frac{2\ell\pi i}{t+1} + \frac{1}{t+1}\log\left(2 + \frac{1.19}{t}\varepsilon_{10}(t,z)\right)\right) \tag{4.14}$$

for some integer $\ell$ with $-\frac{t+1}{2} < \ell \le \frac{t+1}{2}$. In particular, we have

$$\arg z = \frac{2\ell\pi}{t+1} + \frac{1}{t+1}\Im\log\left(1 + \frac{1.19}{2t}\varepsilon_{10}(t,z)\right),$$

which, in view of (4.13), implies $\ell = 0$. Thus (4.14) simplifies to

$$z = \exp\left(\frac{1}{t+1}\log\left(2 + \frac{1.19}{t}\varepsilon_{10}(t,z)\right)\right) = 1 + \frac{\log 2}{t} + \frac{1.63}{t^2}\varepsilon_{11}(t,z).$$

We may now repeat the argument a few times to finally obtain

$$z = 1 + \frac{\log 2}{t} - \frac{\log 2 - \log^2 2}{2t^2} + \frac{4\log^3 2 + 3\log^2 2 + 6\log 2}{24t^3} + \frac{3.45}{t^4}\varepsilon_{12}(t,z).$$

Thus we have $|z| > r_3$. We have therefore shown that

$$|D(q)| \ge \frac{1}{t^3} \qquad \text{for} \qquad |q| = 1/r_3.$$

So we now assume that $D(1/z) = 0$ for some $z$ with $r_3 \le |z| < 1 + \log 2/t$. Repeating the above steps with $1/t^3$ replaced by 0 gives the slightly better bound $z = \rho_2$ with $\rho_2$ as in (2.3).

Inserting the real upper and lower bounds implied by (2.3) into $D_3(q)$ and taking the error $D(q) - D_3(q)$ into account shows that the sign of $D(q)$ changes sign in this interval, so there is a real root $1/z = 1/\rho_2$ of $D(q)$ fulfilling (2.3).

For the $z$ in (2.3), we get

$$D'(1/z) = \frac{2}{\log 2}t^2 + 1.07t\varepsilon_{13}(t,z),$$

which implies that there is exactly one simple zero $1/z$ of $D(q)$ with $z$ fulfilling (2.3). By the same argument as above, this is the only zero $1/z$ with $r_3 \le |z| < 1 + \log 2/t$. Computing $N(1/z)/D'(1/z)$ finally yields the residue given in (2.6).

We already know that $|D(q)| \ge 1/t^3$ for all $q$ with $|q| = 1/r_3$. We also get $|N(q)| \le 5t$. This yields (4.3).

We now turn to the case $2 \le t < 30$. Here, the asymptotic estimates can be replaced by concrete numbers. All assertions have been proved using the interval arithmetic built in in Sage [30]. First, we computed an estimate analogous to (4.11). The corresponding neighbourhood of 2 is subdivided in squares. Each of these squares is intersected with its image under (4.4) and the union of its images under the corresponding analogon to (4.14). If this intersection is empty or the square has no point of absolute value at least $r_3$, the square is discarded. Otherwise, the square is replaced by the smallest square containing the mentioned intersection.

If this does not yield sufficient progress, the squares have been "bisected" into four squares. After a certain number of operations, there are only two small regions which might contain a root. Estimating the derivative, we see that there is at most one root in each of these regions. As it is suspected that these roots are real, the real bisection method is employed to determine the roots with higher precision. The approximation errors $D(q) - D_K(q)$ can also be handled by adding the corresponding interval in the interval arithmetic. ☐

We are now able to prove Theorem 7.

*Proof of Theorem 7.* This is a consequence of singularity analysis [10], cf. also [12].

In this simple case, this also follows from Cauchy's integral formula and the residue theorem (and Proposition 10):

$$\varepsilon_1(t,n)5t^4r_3^n = \frac{1}{2\pi i}\oint_{|q|=1/r_3} \frac{F(q)}{q^{n+1}}\,dq = -R\rho^{n+1} - R_2\rho_2^{n+1} + c_n.$$

☐

## 5. Acknowledgements

C. Heuberger is supported by the Austrian Science Fund (FWF): S09606, that is part of the Austrian National Research Network Analytic Combinatorics and Probabilistic Number Theory. This paper was partly written while C. Heuberger was a visitor at Stellenbosch University.

## References

1. Abedini, N., Khatri, S.P., Savari, S.A.: A SAT-Based Scheme to Determine Optimal Fix-Free Codes, 2010 Data Compression Conference, pages 169–178.
2. Andrews, G. E. The Rogers-Ramanujan reciprocal and Minc's partition function. Pacific J. Math. 95 (1981), no. 2, 251–256.
3. Baer, M.: D-ary Bounded-Length Huffman Coding, ISIT 2007. IEEE International Symposium on Information Theory, 896–900.
4. Bende, S.: The Diophantine equation $\sum_{i=1}^{n} 1/2^{x_i} = 1$ and its connection with graph theory and information theory. (Hungarian) Mat. Lapok 18 (1967), 323–327.
5. Boyd, D.W.: The Asymptotic Number of Solutions of a Diophantine Equation from Coding Theory. 18 (2) (1975), 210–215.
6. Browning, T., Elsholtz, C.: The number of representations of rationals as a sum of unit fractions, Illinois J. Math., to appear.
7. Burkert, J.: Simple bounds on the numbers of binary Huffman sequences. Bull. Inst. Combin. Appl. 58 (2010), 79–82.
8. Chen, Y.G., Elsholtz, C., Jiang, L.-L., Egyptian fractions with restrictions, submitted.
9. Even, S., Lempel, A. Generation and Enumeration of All Solutions of the Characteristic Sum Condition. Information and Control 21, no 5, (1972), 476–482.
10. Flajolet, P., Odlyzko, A.M.: Singularity analysis of generating functions. SIAM J. Discrete Math., vol 3 (1990), 216–240.
11. Flajolet, P., Prodinger, H.: Level number sequences for trees. Discrete Math. 65 (1987), no. 2, 149–156.
12. Flajolet, P., Sedgewick, R., Analytic combinatorics, Cambridge University Press, Cambridge, 2009.
13. Gilbert, E.: Codes based on inaccurate source probabilities, IEEE Transactions on Information Theory 17 (3) (1971), 304–314.
14. Golomb, S.W.: Sources which maximize the choice of a Huffman coding tree, Information and Control, 45 (1980), 263–272.
15. Hoffman, D., Johnson, P., Wilson, N.: Generating Huffman sequences. J. Algorithms 54 (2005), no. 1, 115–121.

16. Huffman, D.A.: A method for the construction of minimum redundancy codes. Proceedings of the IRE 40 9 (1952), pp. 1098–1101.
17. Jackson, B., Ruskey, F.: Meta-Fibonacci sequences, binary trees and extremal compact codes, Electron. J. Combin. 13 (2006), no. 1, Research Paper 26, 13 pp.
18. Khosravifard, M., Esmaeili, M., Saidi H., and Gulliver, T.A., A tree based algorithm for generating all possible binary compact codes with $N$ codewords, IEICE Trans. Fundamentals, Vol. E86-A, No. 10, 2510–2516, Oct. 2003.
19. Khosravifard, M., Saidi, H., Esmaeil, M., Gulliver, T.A.: The Minimum Average Code for Finite Memoryless Monotone Sources, IEEE Trans. Inform. Theory, Vol. 53, No. 3, 955–975, Mar. 2007.
20. Komlos, J., Moser, W., Nemetz, T.: On the asymptotic number of prefix codes, Mitteilungen aus dem Math. Seminar Giessen, Heft 165. Coxeter–Festschrift, Teil III., 1984, 35–48.
21. Lehr, S., Shallit, J., Tromp, J.: On the Vector Space of the Automatic Reals. Theor. Comput. Sci.(1996) 193–210.
22. Minc, H.: A problem in partitions: Enumeration of elements of a given degree in the free commutative entropic cyclic groupoid. Proc. Edinburgh Math. Soc. (2) 11 (1958/1959) 223–224.
23. Narimani, N., Khosravifard, M., The Supertree of the Compact Codes, Proc. Int. Symposium on Telecommunications, Tehran, Iran, 649–655, Aug. 2008.
24. Norwood, E.: The number of different possible compact codes, IEEE Trans. Inform. Theory (1967), pp. 613–616.
25. The On-Line Encyclopedia of Integer Sequences, published electronically at http://oeis.org, 2010.
26. Otter, R.: The Number of Trees. Ann. Math. 49 (1948), 583–599.
27. Parker, D.S., Ram, P.: The Construction of Huffman Codes is a Submodular ("Convex") Optimization Problem Over a Lattice of Binary Trees, SIAM J. Comput. 28 (1999), 1875–1905.
28. Paschke, J., Burkert, J., Fehribach, R.: Computing and estimating the number of $n$-ary Huffman sequences of a specified length, Discrete Mathematics, 311 (1) (2011), Pages 1–7.
29. Savari, S.A.: On Minimum-Redundancy Fix-Free Codes, pp. 3–12, 2009 Data Compression Conference.
30. Stein, W.A. and others, The Sage Development Team, Sage Mathematics Software (Version 4.6.1), `http://www.sagemath.org`, 2011.
31. Tangora, M. Level number sequences of trees and the lambda algebra. European J. Combin. 12 (1991), no. 5, 433–443.

Institut für Mathematik A, Steyrergasse 30/II, Technische Universität Graz, A-8010 Graz, Austria

*E-mail address*: `elsholtz@math.tugraz.at`

Institut für Mathematik B, Steyrergasse 30/II, Technische Universität Graz, A-8010 Graz, Austria

*E-mail address*: `clemens.heuberger@tugraz.at`

Helmut Prodinger, Department of Mathematics, University of Stellenbosch, 7602 Stellenbosch, South Africa

*E-mail address*: `hproding@sun.ac.za`