1987

On the Variance of the External Path Length in a Binary Digital Trie

Peter Kirschenhofer

Helmut Prodinger

Wojciech Szpankowski
Purdue University, spa@cs.purdue.edu

Report Number:

87-680

ON THE VARIANCE OF THE EXTERNAL PATH LENGTH IN A BINARY DIGITAL TRIE

Peter Kirschenhofer
Helmut Prodinger
Wojciech Szpankowski

CSD-TR-680
May 1987

ON THE VARIANCE OF THE EXTERNAL PATH LENGTH IN A BINARY DIGITAL TRIE

(Preliminary version)

Peter Kirschenhofer
Helmut Prodinger
Institut fuer Algebra und
Diskrete Mathematik
TU Vienna
A - 1040 Vienna, Austria

Wojciech Szpankowski
Department of Computer Science
Purdue University
West Lafayette, IN 47907
U.S.A.

Abstract

In this paper we give exact and asymptotic analysis for variance of the external path length in a binary symmetric digital trie. This problem was open up to now. We prove that asymptotically the variance is $4.35\ldots\cdot n + nf(lg\ n)$ where n is the number of stored records and $f(x)$ is a periodic function with a very small amplitude.

1. INTRODUCTION

Digital searching is a well-known technique for storing and retrieving information using lexicographical (digital) structure of words. A *radix trie* (in short: trie) is such a digital search tree that edges are labelled by elements from an alphabet (e.g., binary alphabet consisting of 0's and 1's) and leaves (external nodes) contain keys [1] [3] [7]. More precisely, a key is a (possible infinite) sequence of 0's and 1's, where 0 means "go left" and 1 means "go right". The keys are stored in external nodes and the access path from the root to a leaf is a minimal prefix information contained in an external node. There are a number of applications of tries in computer science and telecommunications, e.g., dynamic hashing, radix exchange sort [3] [7], partial match retrieval of multidimensional data, lexicographical sorting [11], tree-type conflict resolution algorithm for broadcast communications [8] [9], etc.

Two quantities of a digital trie are of special interest: the depth of a leaf and the external path length. The average depth of a leaf has been studied in [2] and [7], the variance in [5] and [9] and higher moments of the depth in [9]. The average value of the external path length is closely related to the average depth of a leaf, but *not* the variance. The variance of the external path length was never determined up to now, although the external path length finds important applications in practice, e.g., for modified lexicographical sorting [11] and for conflict resolution session in conflict resolution algorithms [8]. This paper deals with the exact and asymptotic approximation for the variance of the external path length.

In Section 2, we state the problem to solve and show that the variance of the external path length is associated with a recurrence equation. This equation is solved exactly in Section 3. To find an asymptotic approximation of the solution, we apply either Rice's method or a generalized Mellin transform approach. In fact, these approaches are useful to find an asymptotic approximation for a class of alternative sums. In Section 3, we prove our main result, that is

THEOREM.

The variance of the external path length in a trie consisting of n records (external nodes) is asymptotically equal to

$$\operatorname{var} L_n = n[A + f(lg\ n)] + 0(\ln^2 n) \tag{1.1}$$

where

$$A = 1 + \frac{1}{2\ln 2} - \frac{1}{\ln^2 2} + \frac{2}{\ln 2}(\mu + \nu) + \tau \tag{1.2}$$

$$\mu = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k(2^k - 1)}\ ; \quad \nu = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{2^k - 1} \tag{1.3}$$

$$\tau = -\frac{4\pi^2}{\ln^3 2} \sum_{k=1}^{\infty} \frac{k}{\sinh\left(\frac{2k\ \pi^2}{\ln 2}\right)} = -\frac{1}{4\ln 2} + \frac{1}{\ln^2 2} - \frac{2}{\ln 2} \sum_{j=1}^{\infty} \frac{2^j}{(2^j + 1)^2} \tag{1.4}$$

and $f(x)$ is a continuous periodic function with period 1 and very small amplitude and mean zero.

□

Numerical evaluation reveals that $\operatorname{var} L_n = 4.35 \ldots \cdot n + f(lg\ n)$, where $lg\ n = \log_2 n$.

2. STATEMENT OF THE PROBLEM

Let T_n be a family of tries built from n records with keys from random bit streams. A key consists of 0's and 1's, and we assume that the probability of appearance of 0 and 1 in a stream is equal to p and $q = 1-p$ respectively. The occurrence of these two elements in a bit stream is independent of each other. This defines the so called *Bernoulli model.*

Let L_n denote the external path length (random variable) in T_n, that is, the sum of the lengths of all paths from the root to all external nodes. We are interested in the average value of L_n, $l_n \overset{def}{=} EL_n$, and the variance var L_n. In order to find them, we define the probability generating function, $L_n(z)$ of L_n, that is, $L_n(z) = Ez^{L_n}$. Note that in the Bernoulli model the n records are split randomly into left subtree and right subtree of the root. If X denotes the number of keys in the left subtree, then X is Bernoulli distributed with parameter n and p. Then, for $X=k$, we find

$$E\{z^{L_n} | X = k\} = z^n \, Ez^{L_k} \, Ez^{L_{n-k}} \tag{2.1}$$

where L_k, L_{n-k} represents the external path length in the left subtree (with k keys) and right subtrees ($n-k$ keys). Hence, we obtain

Lemma 1. The probability generating function $L_n(z)$ satisfies the following recurrence

$$\begin{gathered} L_0(z) = L_1(z) = 1 \\ L_n(z) = z^n \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \, L_k(z) L_{n-k}(z) \qquad n \geq 2 \end{gathered} \tag{2.2}$$

□

Let $l_n \overset{def}{=} EL_n$ and $L_n^{''} = EL_n(L_n - 1)$, that is, l_n is the average value of the external path length and $L_n^{''}$ is the second factorial moment of L_n. Note that $l_n = L'(1)$ and $L_n^{''} = L_n^{''}(1)$, where $L_n^{'}(1)$ and $L_n^{''}(1)$ denote the first and the second derivative of $L_n(z)$ at z=1. Simple algebra applied to (2.2) reveals that l_n and $L_n^{''}$ satisfy the following recurrences

$$\begin{gathered} l_0 = l_1 = 0 \\ l_n = n + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (l_k + l_{n-k}) \qquad n \geq 2 \end{gathered} \tag{2.3}$$

and

$$L_0'' = L_1'' = 0$$
$$L_n'' = 2n\ l_n - n(n+1) + 2\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} l_k\ l_{n-k} + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} [L_k'' + L_{n-k}''] \qquad (2.4)$$

Knowing l_n and L_n'' one immediately obtains the variance of L_n, as

$$\text{var}\, L_n = L_n'' + l_n - (l_n)^2 \qquad (2.5)$$

The recurrence (2.4) is a linear one. Hence, let us define three quantities v_n, u_n and w_n as

$$v_0 = v_1 = 0$$
$$v_n = n(n+1) + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (v_k + v_{n-k}) \quad n \ge 2 \qquad (2.6)$$

$$u_0 = u_1 = 0$$
$$u_n = n\ l_n + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (u_k + u_{n-k}) \quad n \ge 2 \qquad (2.7)$$

$$w_0 = w_1 = 0$$
$$w_n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}\ l_k\ l_{n-k} + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} [w_k + w_{n-k}] \quad n \ge 2 \qquad (2.8)$$

Then

$$L_n'' = 2u_n - v_n + 2w_n \qquad (2.9)$$

Note that to compute u_n and w_n we need l_n from recurrence (2.3).

In order to find a uniform approach to solve the recurrence (2.3)–(2.8), we note that all of these recurrences are of the same type and they differ only by the first term which we call the *additive term*. Let, in general, the additive term be denoted by a_n, where a_n is any sequence of numbers. Then the pattern for recurrences (2.3)–(2.8) is

$$x_0 = x_1 = 0$$
$$x_n = a_n + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (x_k + x_{n-k}) \quad n \ge 2 \qquad (2.10)$$

To solve (2.10), let us define a sequence $\hat{a}_n$ (binomial inverse relations [12]) as

$$\hat{a}_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} a_k \iff a_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \hat{a}_k \qquad (2.11)$$

Note that the exponential generating function of $\hat{a}_n$ and a_n satisfies $\hat{A}(-z) = A(z)e^{-z}$. Using this in [9] (see also [7]) it is proved that

Lemma 2. (i) The recurrence (2.10) possesses the following solution

$$x_n = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{\hat{a}_k + ka_1 - a_0}{1 - p^k - q^k} \tag{2.12}$$

(ii) The inverse relatives $\hat{x}_n$ of x_n satisfy

$$\hat{x}_n = \frac{\hat{a}_n + n\, a_1 - a_0}{1 - p^n - q^n} \quad n \geq 2 \tag{2.13}$$

□

Finally, to find asymptotic approximations for x_n, we apply a general approach proposed either in [2] (Rice's method) or in [10] (Mellin like approach, see also Knuth [7]). Namely, we consider an alternative sum of the form $\sum_{k=2}^{n}(-1)^k \binom{n}{k} f(k)$ where $f(k)$ is any sequence. Then

Lemma 3. (i) [Rice's method, see [2], [5]]. Let C be a curve surrounding the points $2, 3, \ldots, n$ and $f(z)$ be an analytical continuation of $f(k)$ in C. Then

$$\sum_{k=2}^{n} \binom{n}{k} (-1)^k f(k) = \frac{-1}{2\pi i} \int_C [n; z] f(z) dz \tag{2.14}$$

with

$$[n; z] = \frac{(-1)^{n-1} n!}{z(z-1) \cdots (z-n)}$$

(ii) [Mellin like approach; see [10]]. Let

$$S_{m,r}(n) = \sum_{k=m}^{n} (-1)^k \binom{n}{k} \binom{k}{r} f(k)$$

and $f(z)$ is an analytical continuation of $f(k)$ left to the line $(½ - [m-r]^+ - i\infty,\ ½ - [m-r]^+ + i\infty)$, $a^+ = \max\{0, a\}$. Then

$$S_{m,r}(n+r) = \frac{(-1)^r}{r!} \int_{(½ - [m-r]^+)} \Gamma(z) f(r-z) n^{r-z} dz + e_n \tag{2.15}$$

where $\int_{(c)}$ stands for $\frac{1}{2\pi i} \int_{c - i\infty}^{c + i\infty}$; $\Gamma(z)$ is the gamma function [4], and

$$e_n = O(n^{-1}) \int_{(1/2 - [m-r])} z\Gamma(z) f(r-z) n^{r-z} \, dz$$

that is, $e_n = o(n)$.

Proof. Both formulas are a consequence of Cauchy's Theorem [4].

□

To apply Lemma 3(i) for asymptotic analysis, we change C to a larger curve around which the integral is small, and take into account residues at poles in the larger enclosed area. To apply 3(ii) we find residues right to the line $(c - i\infty,\ c + i\infty)$

3. EXACT AND ASYMPTOTIC ANALYSIS FOR THE VARIANCE

In this section, we present an exact solution for recurrences (2.3)–(2.8), and asymptotic analysis for the binary *symmetric* case $(p = q = 0.5)$.

Exact analysis

To solve (2.3) for l_n note that $a_n = n$ and $\hat{a}_n = \delta_{n,1}$, where $\delta_{n,k}$ is the Kronecker delta. Then, immediately from Lemma 2 we find

$$l_n = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{k}{1 - p^k - q^k} \tag{3.1}$$

and

$$\hat{l}_n = \frac{n}{1 - p^n - q^n} \qquad n \geq 2 \tag{3.2}$$

To solve (2.6) for v_n, note that $a_n = n(n+1) = 2\binom{n}{2} + 2n$. From [7] [12], we know that for $b_n = \binom{n}{r}$ the inverse relation is $\hat{b}_n = (-1)^r \delta_{n,r}$. Hence, $\hat{a}_n = 2\delta_{n,2} - 2\delta_{n,1}$. By Lemma 2 we obtain

$$v_n = \frac{n(n-1)}{1 - p^2 - q^2} + 2 \sum_{k=2}^{n} (-1)^k \binom{n}{k} \binom{k}{1} \frac{1}{1 - p^k - q^k} \tag{3.3}$$

For u_n given by (2.7), we need the inverse relation for $a_n = n\ l_n$. Using generating functions and the fact $\hat{A}(-z) = A(z)e^{-z}$ one easily proves that $\hat{a}_n = n\ \hat{l}_n - n\ \hat{l}_{n-1}$ where $\hat{l}_n$ is given by (3.2). Hence by Lemma 2

$$u_n = 2\sum_{k=2}^{n}(-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \begin{bmatrix} k \\ 2 \end{bmatrix} \frac{1}{(1-p^k-q^k)^2} + \sum_{k=2}^{n}(-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \begin{bmatrix} k \\ 1 \end{bmatrix} \frac{1}{1-p^k-q^k}$$
$$- 2\sum_{k=3}^{n}(-1)^k \begin{bmatrix} n \\ k \end{bmatrix} \begin{bmatrix} k \\ 2 \end{bmatrix} \frac{1}{(1-p^k-q^k)(1-p^{k-1}-q^{k-1})} \qquad (3.4)$$

The most difficult is w_n since $a_n = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} p^k q^{n-k}\ l_k\ l_{n-k}$. However, let $a(z)$ and $l(z)$ denote the exponential generating functions for a_n and l_n respectively. Then, $a(z) = l(zp)\ l(zq)$, and this implies $\hat{a}(-z) = \hat{l}(-zp)\ \hat{l}(-zq)$. Hence

$$\hat{a}_n = \sum_{k=2}^{n-2} \begin{bmatrix} n \\ k \end{bmatrix} p^k q^{n-k}\ \hat{l}_k\ \hat{l}_{n-k} \qquad n \geq 4 \qquad (3.5)$$

and $\hat{a}_0 = \hat{a}_1 = \hat{a}_2 = \hat{a}_3 = 0$. Then the solution for w_n follows from Lemma 2. We return to that problem later, since (3.5) is not very suitable for analytical continuation needed in Lemma 3.

Asymptotic approximation

In this preliminary report, we present the asymptotic analysis only for binary *symmetric* tries, that is, hereafter we assume $p = q = 0.5$. We obtain asymptotic approximations for v_n and w_n, through Lemma 3(ii) and for w_n by Lemma 3(i), however, both methods are equivalent.

Let us start with v_n. Note that $v_n = 2l_n + 2n^2 - 2n$. Using the asymptotic expression for l_n from [5] [7], we immediately obtain (see also (3.24))

$$v_n = 2n^2 + \frac{2n\ \ln n}{L} + n\left[\frac{2\gamma}{L} - 1\right] + 2n\ \delta(lg\ n) \qquad (3.6)$$

where $\gamma = 0.577...$ is the Euler constant, $L \overset{def}{=} \ln 2$, and

$$\delta(x) = \frac{1}{L}\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \omega_k \Gamma(-\omega_k) e^{2k\pi i x} \qquad (3.7)$$

where

$$\omega_k = 1 + \frac{2k\pi i}{\ln 2} \qquad (3.8)$$

The ω_k, $k = 0, \pm 1, \ldots,$ are solutions of the following equation

$$1 - 2^{1-z} = 0 \tag{3.9}$$

where z is a complex number.

The evaluation of u_n is much more intricate. Using (3.4) with $p = q = 0.5$ one proves

$$\frac{u_{n+1}}{n+1} = 8n + \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{1}{(1-2^{-k})^2} \left[\frac{k}{2(2^{k-1}-1)} - 1 \right]$$

Hence by Lemma 3(ii)

$$\frac{u_{n+1}}{n+1} - 8n = \int_{(-3/2)} \frac{\Gamma(z) n^{-z}}{(1-2^z)^2} \left[\frac{-z}{2(2^{-z-1}-1)} - 1 \right] dz + 0(n^{-1}) \tag{3.10}$$

The evaluation of the integral is standard and appeals to the residue theorem. Note that the function under the integral has two poles : $-\omega_k$ given by (3.8) and $\chi_k = 2k\pi i/L$ for $k = 0, \pm 1, \pm 2, \ldots,$ $(\omega_k = 1 + \chi_k)$. For k=0, $-\omega_0 = -1$ is a double pole, while $\chi_0 = 0$ is a triple pole since z=0 and z=−1 are a singular points for $\Gamma(z)$. It is also well known that the main contribution to the asymptotic approximation comes from the real part of the poles, that is, $-\omega_0$ and χ_0. For $-\omega_0 = -1$ we use the following Taylor expansions. Let $w = z+1$, then

$$\Gamma(z) = -w^{-1} - (\gamma - 1) + 0(w) \tag{3.11a}$$

$$n^{-z} = n(1 - w \ln n) + 0(w^2) \tag{3.11b}$$

$$\frac{1}{2^{-z-1} - 1} = \frac{-1}{Lw} \left\{ 1 + \frac{Lw}{2} + \frac{L^2 w^2}{12} \right] \tag{3.11c}$$

For $\chi_0 = 0$ we have [4]

$$\Gamma(z) = z^{-1} + \gamma + ½ \left[\frac{\pi^2}{6} + \gamma^2 \right] z + 0(z^2) \tag{3.12a}$$

$$n^{-z} = 1 - z \ln n + \frac{z^2}{2} \ln^2 n + 0(z^3) \tag{3.12b}$$

$$\frac{1}{2(2^{-z-1} - 1)} = -(1 + Lz) \tag{3.12c}$$

$$\frac{1}{(1-2^{-z})^2} = \frac{1}{L^2 z^2}\,(1 + Lz + \frac{5}{12}L^2 z^2) \tag{3.12d}$$

Multiplying (3.11) and (3.12) and taking the coefficient at z^{-1} and w^{-1} we find the contribution from $-\omega_0$ and χ_0 which yields

$$u_{n,1} = \frac{n}{L}\left[2n \,\ln n + n(2\gamma - L) + \frac{1}{2L}\ln^2 n + \left(\frac{\gamma+1}{L} - 1\right)\ln n \right. \\ \left. + \left(\frac{\gamma}{L} + \frac{\pi^2}{12L} + \frac{\gamma^2}{2L} + \frac{17}{12}L - 1 - \gamma\right)\right] \tag{3.13}$$

The contribution from $-\omega_k$ and χ_k, $k \neq 0$ can be found in a similar way. Calculations reveal that

$$u_{n,2} = \frac{n^2}{L}\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} \exp[2k\pi i \; lg \; n]\left\{ 2\,[\omega_k\Gamma(-\omega_k) + \frac{c_1\omega_k}{n} - \frac{\omega_k^2\Gamma(-\omega_k)}{n}\,] + \right. \\ (-\omega_k)[(-\omega_k)\Gamma(-\omega_k)\ln n + \omega_k\Gamma'(-\omega_k) - \Gamma(-\omega_k) \\ \left. - L\Gamma(-\omega_k) - \Gamma(-\omega_k)(1 + 2L\chi_k)] \right\} \tag{3.14}$$

$$c_1 = -\chi_k\omega_k\Gamma(-\omega_k)/2 \tag{3.15}$$

Finally, $u_n \sim u_{n,1} + u_{n,2}$ and it turns out that the contribution from $u_{n,2}$ is very small.

The most difficult part is the asymptotic approximation for w_n, since we need analytical continuation for $\hat{a}_n$ given by (3.5). Fortunately for the symmetric case, it is relatively easy to obtain, however, we need a further consideration to find it. Note that for $a_n = 2^{-n}\sum_{k=0}^{n}\binom{n}{k} l_k l_{n-k}$ the exponential generating function $a(z)$ and $l(z)$ are related as $a(z) = [l(z/2)]^2$, and $\hat{a}(z) = [\hat{l}(z/2)]^2$. From (2.3) with $p = q = 0.5$ we immediately find

$$\hat{l}(z) = z(e^z - 1) + 2\hat{l}(z/2) \tag{3.16}$$

Hence

$$[\hat{l}(z)]^2 = z^2(e^{2z} - 2e^z + 1) + 4z(e^z - 1)\hat{l}(z/2) + [\hat{l}(z/2)]^2$$

and equating coefficients of both sides of the above, we finally obtain for $n \geq 3$ (note that $\hat{l}^2(z) = \hat{a}(2z)$)

$$\hat{a}_n = \frac{n(n-1)}{2} \frac{1}{2^{n-2}-1}\left[2^{n-3} - 1 + \sum_{j=1}^{\infty} \binom{n-2}{j} \frac{1}{2^j - 1} - \frac{1}{2^{n-2}-1}\right]$$

Hence, by Lemma 2 w_n has a solution

$$\frac{w_{n+1}}{n+1} = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{2^{k-1}}{2^k - 1} \frac{k}{2^{k-1}-1}\left\{1 - 2^{k-2} + \frac{1}{2^{k-1}-1} - \sum_{j=1}^{\infty}\binom{n-1}{j}\frac{1}{2^j - 1}\right\} \quad (3.17)$$

For asymptotic analysis of (3.17), we apply Rice's method to illustrate how it works. Note that the analytical continuation of $f(k)$ in (3.17) is easy, since the last series in (3.17) can be extended as $\sum_{j=1}^{\infty}\binom{z-1}{j}\frac{1}{2^j-1}$. Hence, using Rice's method

$$\frac{w_{n+1}}{n+1} = -\frac{1}{2\pi i}\int_C [n\,;z] f(z)dz$$

where

$$f(z) = \frac{z\,2^{z-1}}{(2^z - 1)(2^{z-1}-1)}\left\{1 - 2^{z-2} + \frac{1}{2^{z-1}-1} - \sum_{j=1}^{\infty}\binom{z-1}{j}\frac{1}{2^j-1}\right\} \quad (3.18)$$

We extend now the circle of the integration such that the poles of $f(z)$ are included, that is, the points ω_k and χ_k, $k = 0 \pm 1, \ldots,$. We evaluate separately the residues of the function under the integral for $\omega_0 = 1$, $\chi_0 = 0$ and $\omega_k, \chi_k, k{\neq}0$. We use the Taylor expansions already presented in (3.11) and (3.12). In addition, we have for $w = z-1$ (see [2] [4] [6])

$$[n\,;z] \sim \frac{n}{w}[1 + w(H_{n-1} - 1) + w^2(1 - H_{n-1} + ½\, H_{n-1}^2 + ½\, H_{n-1}^{(2)})] \quad (3.19a)$$

$$\sum_{j=1}^{\infty}\binom{z-1}{j}\frac{1}{2^j-1} \sim \mu + 0(w) \quad \text{for } z \to 1 \quad (3.19b)$$

$$\sum_{j=1}^{\infty}\binom{z-1}{j}\frac{1}{2^j-1} \sim \nu + 0(z) \quad \text{for } z \to 0 \quad (3.19c)$$

where H_n, $H_n^{(2)}$ are harmonic and generalized harmonic numbers, μ and ν are defined in (1.3). Then the contributions from $\omega_0 = 1$ and $\chi_0 = 0$ to w_n denoted as $w_{n,1}$, $w_{n,2}$ are respectively

$$w_{n,1} = \frac{n}{L^2}\left\{½\, n \ln^2 n + \gamma n \ln n - \frac{3L}{2} n \ln n + n\alpha - ½ \ln^2 n \right.$$
$$\left. + \left(\frac{3}{2}L - \gamma - \frac{3}{2}\right)\ln n - \alpha + \frac{9}{4}L - \frac{3\gamma}{2} - ½\right\} \quad (3.20)$$

where $\alpha = \frac{5}{3}L^2 - \frac{3L\gamma}{2} - L\mu + \frac{\gamma^2}{2} + \frac{\pi^2}{12}$, and

$$w_{n,2} = \frac{n}{L}\left\{-\frac{5}{4} + \nu\right\} \tag{3.21}$$

To find the contribution from ω_k and χ_k, $k \neq 0$, we use the following Taylor expansions for $u = z - \omega_k$

$$[n\,;z] \sim n^{\omega_k}\{\Gamma(-\omega_k) + u[-\Gamma'(-\omega_k) + \Gamma(-\omega_k)\ln n]\}$$

Then the contribution, $w_{n,3}$ from ω_k, $k \neq 0$ is

$$\begin{aligned} w_{n,3} = \frac{n^2 \ln n}{L}\,\delta(lg\ n) + \frac{n^2}{L^2}\,\sigma_1(lg\ n) + \frac{n\ln n}{L^2}\,\sigma_2(lg\ n) \\ + \frac{n}{L^2}\,\sigma_3(lg\ n) \end{aligned} \tag{3.22}$$

where $\delta(x)$ is defined in (3.7) while $\sigma_i(x)$, $i = 1, 2, 3$ are complicated fluctuating functions with very small amplitude (see also (3.27b)). Finally, the contribution from χ_k, $k \neq 0$ is

$$w_{n,4} \sim \frac{1}{L}\sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} n^{\omega_k}\Gamma(-\omega_k)\omega_k\,\chi_k\left[-5/4 - \sum_{j=1}^{\infty}\binom{\chi_k - 1}{j}\frac{1}{2^j - 1}\right] \tag{3.23}$$

and $w_n = w_{n,1} + w_{n,2} + w_{n,3} + w_{n,4} + 0(ln^2 n)$.

To complete our analysis, we need the asymptotic approximation for l_n. But from [5] [9] we have

$$l_n \sim \frac{n\ln n}{L} + n\left[\frac{\gamma}{L} + ½ + \delta(lg\ n)\right] - \frac{1}{2L} + \delta_1(lg\ n) \tag{3.24}$$

where $\delta(x)$ is defined in (3.7) and

$$\delta_1(x) = -\frac{1}{L}\sum_{k\neq 0}\frac{\omega_k^2\chi_k}{2}\,\Gamma(-\omega_k)e^{2\pi kix} \tag{3.25}$$

Now, the variance of L_n is given by var $L_n = 2u_n - v_n + 2w_n + l_n - l_n^2$, and after some tedious algebra, one finds

$$\text{var}\ L_n = B\,n^2 + A\,n + 0(\ln^2 n) \tag{3.26}$$

where

$$B = -\frac{11}{12} - \frac{2\mu}{L} + \frac{\pi^2}{6L^2} - \delta^2(lg\ n) + \left[3 - \frac{2\gamma}{L}\right]\delta(lgn) + \frac{2}{L^2}\,\sigma_1(lg\ n) \qquad (3.27a)$$

and

$$\sigma_1(x) = \sum_{\substack{k=-\infty \\ k\neq 0}}^{\infty} e^{2\pi kix}\left\{\Gamma(-\omega_k) - \frac{3L}{2}\,\omega_k\Gamma(-\omega_k) - \omega_k\Gamma'(-\omega_k) - \right.$$

$$\left. L\,\omega_k\Gamma(-\omega_k)\sum_{n\geq 1}\begin{bmatrix}\chi_k \\ n\end{bmatrix}\frac{1}{2^n-1}\right\} \qquad (3.27b)$$

and A is given in the main Theorem (see (1.2)).

To prove the main Theorem, we need to show that $B=0$. Let us first consider the Fourier coefficient of $\delta^2(x)$ for $k=0$. We denote it by δ_0. Then

$$\delta_0 = \frac{1}{L^2}\sum_{\substack{l+m=0 \\ l,\,m\neq 0}} \omega_l\,\omega_m\ \Gamma(-\omega_k)\Gamma(-\omega_m) =$$

$$\frac{2}{L^2}\sum_{l=1}^{\infty}\Gamma(\chi_l)\,\overline{\Gamma(\chi_l)} = \frac{1}{L}\sum_{l=1}^{\infty}\frac{1}{l\ \sinh\,(2l\xi)} \qquad (3.28)$$

where $\xi = \pi^2/L$. This can be rewritten as

$$\delta_0 = 2\ \ln\prod_{n=1}^{\infty}\frac{1}{1-e^{-2\xi n}} - 2\ \ln\prod_{n=1}^{\infty}\frac{1}{1-e^{-4\xi n}}$$

Using the functional equation for the Dedekind η-function Kirschenhofer, Prodinger and Schoissengeier [6] have proved that δ_0 can be reduced to

$$\delta_0 = \frac{\pi^2}{6L} - L + \frac{L}{12} - 2\mu$$

Hence the constant B in (3.27) can be transformed into

$$B = -\delta_2^2(lg\ n) + [3 - \frac{2\gamma}{L}]\delta(lgn) + \frac{2}{L^2}\,\sigma_1(lg\ n) = \delta_3(lg\ n)$$

where $\delta_2^2(x)$ is the function $\delta^2(x) - \delta_0$. Note that now B is expressed in terms of a periodic function $\delta_3(x)$, with very small amplitude and mean zero. This function is continuous (since the Fourier series associated with the function is absolutely convergent). Assume now $\delta_3(x)$ is not identically zero. Then, $\delta_3(x)$ would take values, say less that $-\varepsilon$, for arguments in an interval, say [a,b]. Since $\lg_2 n$ is dense modulo 1, the leading factor of the variance would be negative for infinitely many values of n. This is a contradiction, since var $L_n \geq 0$ for all n. Hence $\delta_3(x) \equiv 0$ and thus $B = 0$. This completes the proof of the main Theorem.

References

1. **Aho, A., Hopcroft, J.** and **Ullman, J.**, Data structures and algorithms, Addison-Wesley (1983).

2. **Flajolet, Ph.** and **Sedgewick, R**, Digital search trees revisited, *SIAM J. Comput.*, 15, pp. 748–767 (1986).

3. **Gonnet, G.**, Handbook of algorithms and data structures, Addison-Wesley (1986).

4. **Henrici, P.**, Applied and computational complex analysis, John Wiley & Sons, New York (1977).

5. **Kirschenhofer, P.** and **Prodinger, H.**, Some further results on digital trees, in : *iAutomata, Languages and Machines* (L. Kott ed.) , 177 - 185, Springer Lectures Notes in Computer Science, 226, (1986).

6. **Kirschenhofer, P., Prodinger, H.** and **Schoissengeier, J.**, Zur Auswertung gewisser Reihen mit Hilfe modularer Funktionen, to appear in : *Zahlentheortische Analysis* 2 (E. Hlawka ed.), Springer Lecture Notes in Math.

7. **Knuth, D.**, The art of computer programming. Sorting and searching. Addison-Wesley (1973).

8. **Mathys, P., Flajolet, P.**, Q-ary collision resolution algorithms in random-access system with free and blocked channel access, *IEEE Trans. Information Theory*, vol. IT–31, 2, pp. 217–243 (1985).

9. **Szpankowski, W.**, Some results on V-ary asymmetric tries, Purdue University, CSD-TR 582 (1986) (submitted to a journal).

10. **Szpankowski, W.**, Two problems on the average complexity of digital trees, submitted to a conference.

11. **Paige, R.** and **Tarjan, R.**, Three efficient algorithms based on partition refinement, preprint (1986).

12. **Riordan, J.**, Combinatorial identities, John Wiley & Sons (1968).