1991

Multidimensional Digital Searching and Some New Parameters in Tries

Peter Kirschenhofer

Helmut Prodinger

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

Report Number:

91-052

MULTIDIMENSIONAL DIGITAL SEARCHING
AND SOME NEW PARAMETERS IN TRIES

Peter Kirschenhofer
Helmut Prodinger
Wojciech Szpankowski

CSD-TR-91-052
July 1991

MULTIDIMENSIONAL DIGITAL SEARCHING AND SOME NEW PARAMETERS IN TRIES

June 29, 1991

Peter Kirschenhofer*
Dept. of Algebra&Discrete Math.
Technical University of Vienna
Vienna A-1040
Austria

Helmut Prodinger†
Dept. of Algebra&Discrete Math.
Technical University of Vienna
Vienna A-1040
Austria

Wojciech Szpankowski‡
Dept. of Computer Science
Purdue University
W. Lafayette, IN 47907
U.S.A.

Abstract

Multidimensional digital searchings (M-d tries) are analyzed from the view point of partial match retrieval. Our first result extends the analysis of Flajolet and Puech of the average cost of retrieval to biased probabilities of symbols occurrences in a key. The second main finding concerns the variance of the cost of the retrieval. This variance is of order $O(N^{1-s/M})$ where N is the number of records stored in a M-d trie, and s is the number of specified components in a query of size M. For $M = 2$ and $s = 1$ we present a detailed analysis of the variance, which identifies the constant at $\sqrt{N}$. This analysis, which is the central part of our paper, requires certain series transformation identities which go back to Ramanujan. In the Appendix we provide a Mellin transform approach to these results. Finally, we point out that the cost of the partial match retrieval can be used to analyze some other phenomena on strings such as long runs of repetitive patterns, etc.

*†This work is supported by Fonds zur Förderung der Wissenschaftlichen Forschung Project P7497-TEC
‡This research was supported by the NSF Grant INT-8912631, and in part by AFOSR grant 90-0107, by the NSF grant CCR-8900305, and by grant R01 LM05118 from the National Library of Medicine.

1. INTRODUCTION

Multidimensional searching is of prime interest to a nowadays computer science. In particular, retrieval of multidimensional data found applications in the design of data base systems and graphics. Bentley [7] and Rivest [26] are founding fathers of multidimensional searching, namely digital M-d trees and M-d search trees. An early description of these structures can be found in volume three of Knuth's book [23], and a more detailed discussion is in the book of Mahmoud [24]. Several applications of multidimensional searching are discussed in the paper of Flajolet and Puech [13], which also presents a thorough analysis of partial matching retrieval for multidimensional data.

In this paper we concentrate on M-d digital trees (i.e., *M-multidimensional tries*) and partial retrieval in such a data structure. A brief description of tries follows (cf. also [5], [23], [21] and [22], [24]). In general, a trie stores a set of, say N, keys (words, strings) from a set Σ^*, which is a set of all (possible infinite) sequences built over a finite alphabet Σ. A trie is composed of branching nodes, called also *internal nodes*, and *external nodes* that store the keys. In addition, we assume that every external node is able to store only one key. The branching policy at any level, say k, is based on the k-th symbol of a key. For example, for a binary alphabet $\Sigma = \{0, 1\}$, if the k-th symbol in a key is "0", then we branch-out left in the trie, otherwise we go to the right. This process terminates when for the first time we encounter a different symbol between a key that is currently inserted into the trie and all other keys already in the trie. Then, this new key is stored in a newly generated external node. In other words, the access path from the root to an external node (a leaf of a trie) is the minimal prefix of the information contained in this external node; it is minimal in the sense that this prefix is not a prefix of any other keys.

The M-d tries are built in a similar manner, however, this time a key is an M-dimensional tuple. Hereafter, we only consider the binary alphabet $\Sigma = \{0, 1\}$. Then, the ith key has a structure as follows: $K_i = (K_{i1}, \ldots, K_{iM})$, where $K_{ij} \in \Sigma^*$ and $1 \le i \le N$. In an M-d trie subkeys K_{i1}, ... ,K_{iM} are "shuffled" to yield one binary sequence $\tilde{K}_i$, and these new composite keys, $\tilde{K}_1, \ldots, \tilde{K}_N$, are used to construct a regular trie. More precisely, let $K_i = (K_{i1}, \ldots, K_{iM})$ and $K_{i\ell} = (k_{i\ell}^1, k_{i\ell}^2, ...)$ where $k_{i\ell}^m \in \Sigma = \{0, 1\}$ for $1 \le \ell \le M$ and $m \ge 1$. Then, the new ith key is $\tilde{K}_i = (k_{i1}^1 k_{i2}^1 \cdots k_{iM}^1 k_{i1}^2 k_{i2}^2 \cdots k_{iM} \cdots)$. In passing, we note that the statistics of the composite keys are the same as the original keys provided the keys are statistically independent.

A *partial match query* q requests all records that match the query. There are two types of query components: unspecified denoted by $*$ and specified component $S \in \Sigma^*$. For example, $q = (S_1, *, S_2)$ where $S_1, S_2 \in \Sigma^*$ has two specified components S_1, S_2 and one unspecified one. By the shuffling procedure, as described above, we transform the query q into a *search pattern* $\omega \in \Sigma^*$, that is, the first symbol of ω comes from the first symbol of the first component of q (if the first component is unspecified $*$, then we put $*$ in ω too), the second symbol of ω originates from the first symbol of the second component of q, and so on modulo M. The search pattern is used to search for external nodes in a M-d trie that satisfy the pattern, where "1" in ω means move right, "0" means move left, and "$*$" means to proceed in both directions. The *cost of a query* is defined as the number of internal nodes visited during such a search.

The cost of query depends on the search pattern ω. In fact, the search pattern specifies a subtree in a trie, and can be considered as a new parameter that characterizes the trie. For example, if a query consists only of specified patterns, then the query defines a specific path (depth) in the trie. This path may be of various shapes depending on the search pattern. Indeed, if $\omega = (11111111\cdots)$, then the query finds the right-most internal nodes, and the cost of the query is the length of such a path; the length of a "zig-zag" path can be computed by evaluating the cost of the following search pattern $\omega = (010101\cdots)$, etc. Moreover, such a partial match query can be used to identify the *longest prefix of long repetitive patterns* (e.g., long run of 1's) common to at least two out of a set of stored keys (strings). For example, if the repetitive pattern is B, then the search pattern should be defined as $\omega = (BBBBB\cdots)$. On the other hand, if the query consists of only unspecified components $*$, then the cost of the query is equivalent to the size of the underlying trie. Finally, with a mixture of specified and unspecified components, the cost of a query represents the length of a subtree of the trie.

The partial match retrieval has some potential applications in algorithms for *pattern matching with errors*. In this case, a trie is replaced by a suffix tree (cf. [3] and [5]) of a the *text* word. The *pattern* word is of length M, and it is assumed that some (known) positions may contain an error. Then, to find all occurrences of the pattern in the text word – allowing errors to occur in the pattern – one needs to search for a partial query with unspecified components in the positions of potential errors. Such algorithms are of prime interest to biologists, but also some other applications can be envisioned (e.g., errorness transmissions, etc.). Finally, using suffix tree and partial match query we can easily find the longest run of a given pattern B in a *single* random sequence (cf. Erdös and Révész [12], and Guibas and Odlyzko [16]). For example, for finding the longest run of 1's, we need to search for $\omega = (1111\cdots)$, while the longest run of $B = (0,1)$ we have $\omega = (01010101\cdots)$.

The literature on partial matching and multidimensional searching is huge (cf. [7],

[13], [23], [26]). However, it is lack of detailed analyses. The gap was partially fulfilled by a prominent paper of Flajolet and Puech [13]. The authors of [13] – using ingenious techniques from complex analysis and differential equations – presented a very detailed evaluation of the *average* cost of a query. This analysis was done in a probabilistic framework which assumed equal probabilities for symbols occurrences.

In this paper we extend Flajolet and Puech's results in two directions. First of all, we complete the study initiated by the authors of [13], and present the average cost of a query in the asymmetric case, that is, when probabilities of symbols occurrences are not identical. More importantly, we also analyze the variance of the cost for the special case $M = 2$, which already requires rather sophisticated mathematical techniques. In this case we shall show that the variance of the cost is of order $O(\sqrt{N})$, and this will imply the convergence *in probability* of the cost to its average value. We also conjecture that for general M the variance is of order $O(N^{1-s/M})$ where s is the number of specified components in a query. We point out that this can be difficult to prove, especially if one needs to identify the constant at $N^{1-s/M}$.

This paper is organized as follows. In the next section we present our main results, and discuss some consequences of them. All proofs are delayed till Section 3. The asymptotic analysis of Section 3 resembles the one in [13]. It is based on solving some recurrences equation through generating function approach, and obtaining asymptotics of some alternating sums through the usage of Rice's formula (cf. [14]) or the Mellin transform approach (cf. [27]). The main difficulty in this analysis lies in proving that some coefficients in the asymptotics of the variance are equal to zero. These kind of problems were already tackled by us in [21] and [22]. However, this time the difficulty is one step higher. More precisely, during the course of our analysis we require transformation results for $F(x) = \sum_{k\geq 1} e^{-kx}/(1+e^{-2kx})$ and related functions. It turns out that by a very general formula of Ramanujan (cf. [8]) we have $F(x) = \pi/4x - 1/4 + \pi F(\pi^2/x)/x$ for $x > 0$. In the Appendix we give an outline of a Mellin transform proof for this and similar transformation results.

2. MAIN RESULTS

In this section we present our main results concerning the cost of a query. For simplicity of presentation we discuss only the binary alphabet $\Sigma = \{0, 1\}$. We adopt the following probabilistic model: *symbols from the alphabet are generated independently, and the ith symbol occurs with probability p_i.* Our second main result – concerning the variance of the cost – is restricted to the symmetric alphabet, that is, for $p_0 = p_1 = 0.5$

Let q be a query and ω a query search pattern, as discussed in the Introduction. We extend the alphabet to include the unspecified query, that is, $\tilde{\Sigma} = \{0, 1, *\}$ and the set of all infinite sequences of $\tilde{\Sigma}$ is denoted as $\tilde{\Sigma}^*$. Note that $\omega \in \tilde{\Sigma}^*$. Moreover, for M-dimensional queries the following property of the query search pattern holds. Let $\omega^{(k)}$ denote a shift to the left by k positions of ω. Then by definition $\omega = \omega^{(M)}$, and for simplicity of notation we shall describe ω by specifying only the first M symbols (e.g., we write $\omega = (*, S)$ instead of the correct $\omega = (*, S, *, S, \cdots)$). We also write for simplicity $\omega' = \omega^{(1)}$. We use this fact to derive the generating function of the cost.

Let $H_N^\omega(z)$ be the probability generating function of the cost of the query ω. To present our analysis in a compact form, we need a little bit of notation. Define a set of zero-one M-tuples by Θ_M, that is, $\Theta_M = \{\mathrm{b} = (b_1, \cdots, b_M) : b_i \in \{0, 1\}\}$. We also generalize the Kronecker delta function. For $w \in \{0, 1, *\}$ and $b \in \{0, 1\}$, we define

$$\delta(b, w) = \begin{cases} 1 & \text{if } w = * \\ \delta_{b,w} & \text{if } w \in \{0, 1\} \end{cases} \tag{1}$$

where $\delta_{b,w} = 1$ when $b = w$ and zero otherwise. Using this notation, and generalizing the approach of Flajolet and Puech [13] (cf. also [21] and [22]) we obtain the following recurrence equation for the probability generating function for $N \geq 2$

$$H_N^\omega(z) = \begin{cases} z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_k^{\omega'}(z) H_{N-k}^{\omega'}(z) & \omega = *\omega' \\ z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_k^{\omega'}(z) & \omega = 0\omega' \\ z \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} H_{N-k}^{\omega'}(z) & \omega = 1\omega' \ , \end{cases} \tag{2}$$

and $H_0^\omega(z) = H_1^\omega(z) = 1$.

The average cost of the query ω is simply the first derivative of $H_N^\omega(z)$ at $z = 1$. Let it be denoted as l_N^ω. Using our generalized Kronecker delta notation (1), after some algebra one proves the following recurrence for l_N^ω.

$$l_N^\omega = 1 + \sum_{k=0}^N \binom{N}{k} p_0^k p_1^{N-k} \{\delta(0, \omega_1) l_k^{\omega'} + \delta(1, \omega_1) l_{N-k}^{\omega'}\} \ , \tag{3}$$

and $l_0^\omega = l_1^\omega = 1$. This recurrence can be solved as in [13], [14], and in [21], [22]. In Section 3 we present more details how to handle (3).

Now we are ready to present our first result concerning the average cost of the query ω.

Theorem 1. (i) *Let $s = M$ (only specified components in a query), and the search pattern ω is a repetitive pattern with period M. There are m zeros and $M - m$ ones in the repetitive* *pattern. Then, for large N*

$$l_N^\omega = M\frac{\log N}{\log\left(p_0^{-m}p_1^{m-M}\right)} + O(1)\ . \tag{4}$$

(ii) *Let $0 < s < M$. Define $-1 < \kappa_0 < 0$ as the unique solution of the following equation*

$$\sum_{\mathbf{b}\in\Theta_M}\prod_{i=1}^{M}\delta(b_i,\omega_i)p_{b_i}^{-\kappa_0} = 1\ . \tag{5}$$

Then, the average cost becomes for large N

$$l_N^\omega = N^{-\kappa_0}\left(\frac{\Gamma(\kappa_0)(1+\kappa_0)\sum_{j=1}^{M-1}\sum_{\mathbf{b}\in\Theta_j}\prod_{i=1}^{j}\delta(b_i,\omega_i)p_{b_i}^{-\kappa_0}}{\sum_{\mathbf{b}\in\Theta_M}\log\left(\prod_{1\le i\le M}^{*}\delta(b_i,\omega_i)p_{b_i}\right)} + \xi(N)\right) + O(1) \tag{6}$$

where $\xi(N)$ is an oscillating function with small amplitude, and the product $\prod_{1\le i\le M}^{*}$ is taken only over nonzero terms. ■

For example, for $M = 2$ and $\omega = (*,1)$ equation (5) becomes

$$p_1^{-\kappa_0}(p_1^{-\kappa_0} + p_0^{-\kappa_0}) = 1\ ,$$

and then the cost of the query is

$$l_N^\omega = N^{-\kappa_0}\left(\frac{\Gamma(\kappa_0)(1+\kappa_0)(p_1^{-\kappa_0}+p_0^{-\kappa_0})}{\log p_0p_1 + \log p_1^2} + \xi(N)\right) + O(1)\ .$$

The above should be compared with the cost (20) of l_N^ω with $\omega = (*,S)$ in the symmetric case.

Our next result concerns the variance of the cost in the case $0 < s < M$ (i.e., at least one unspecified component of a query). This analysis is much more intricate, as demonstrated in our papers [21], [22] that are devoted to the variance of the external path length in a trie and PATRICIA respectively. Therefore, hereafter we assume only the symmetric alphabet, that is, "0" and "1" occur with equal probability 0.5. In such a case, the recurrence equation (2) for the probability generating function becomes

$$H_N^\omega(z) = \begin{cases} z\sum_{k=0}^{N}\binom{N}{k}2^{-N}H_k^{\omega'}(z)H_{N-k}^{\omega'}(z) & \omega = *\omega' \\ z\sum_{k=0}^{N}\binom{N}{k}2^{-N}H_k^{\omega'}(z) & \omega = S\omega' \end{cases}, \tag{7}$$

where $\omega' = \omega^{(1)}$.

If the cost of the query ω is denoted by C_N^ω, then the variance var C_N^ω can be evaluated according to the following formula

$$\text{var}\ C_N^\omega = \frac{d^2}{dz^2}H_N^\omega(z)\mid_{z=1} + l_N^\omega - (l_N^\omega)^2\ , \tag{8}$$

where the second derivative of the generating function at $z = 1$ represents the second factorial moment of the cost, that is, $EC_N^\omega(C_N^\omega - 1)$.

The evaluation of the second factorial moment of C_N^ω leads to very tedious algebraic manipulations, although the main idea behind it is not too complicated. In the following we restrict our investigations to the case of $M = 2$, i.e., $\omega_1 = (*, S)$ and $\omega_2 = (S, *)$. As shown in Section 3, the analysis of the centralized moments in this case is a rather sophisticated one.

Theorem 2. (i) *For $\omega_1 = (*, S)$ we have asymptotically*

$$\text{var } C_N^{\omega_1} = \sqrt{N}(A_1 + \tau_1(\log_2 \sqrt{N})) + o(\sqrt{N}) \simeq 2.0918454253\sqrt{N} \ , \tag{9}$$

where the constant A_1 can be expressed analytically as

$$\begin{aligned} A_1 &= \frac{7\sqrt{\pi}(1+\sqrt{2})}{2L} - 2\sqrt{\pi}\left(\frac{1}{L}\left[\frac{73+25\sqrt{2}}{16\sqrt{2}} - \frac{4}{3\sqrt{3}}\right]\right. \\ &+ \left.\frac{2}{L}\sum_{\ell\geq 2}\binom{1/2}{\ell}\frac{(\ell-1)(\ell+\frac{1}{2})(1+2^{-\ell})(1+2^{\frac{1}{2}-\ell})2^{-\ell}}{1-2^{1-2\ell}}\right) \ , \end{aligned} \tag{10}$$

with $L = \log 2$, and $\tau_1(x)$ is a continuous periodic function of period 1 and very small amplitude.
For $\omega_2 = (S, *)$ we have

$$\text{var } C_N^{\omega_1} = \sqrt{N}(A_2 + \tau_2(\log_2 \sqrt{N})) + o(\sqrt{N}) \simeq 1.1532422249\sqrt{N} \ , \tag{11}$$

where the constant A_2 becomes

$$\begin{aligned} A_2 &= \frac{5\sqrt{\pi}(1+\sqrt{2})}{2\sqrt{2}L} - 2\sqrt{\pi}\left(\frac{1}{4L}\left[\frac{57+\sqrt{2}}{8} + 2\sqrt{2} - \frac{8}{3}\sqrt{\frac{2}{3}}\right]\right. \\ &+ \left.\frac{\sqrt{2}}{L}\sum_{\ell\geq 2}\binom{1/2}{\ell}\frac{(\ell-1)(\ell-\frac{1}{2})(1+2^{-\ell})(1+2^{\frac{1}{2}-\ell})2^{-\ell}}{1-2^{1-2\ell}}\right) \ , \end{aligned} \tag{12}$$

and $\tau_2(x)$ is again of period 1.

(ii) *The cost of the query ω converges in probability to $EC_N^\omega = \ell_N^\omega$, that is, the following holds*

$$\lim_{N\to\infty} \Pr\{\mid C_N^\omega / l_N^\omega - 1 \mid > \varepsilon\} = 0 \tag{13}$$

for any $\varepsilon > 0$. ■

Remarks.

(i) *Repetitive Patterns.* Our analysis of the partial match query leads to some insights into problems on strings. In particular, consider the following problem. Let B be a pattern of length M, and we ask for the longest prefix of at least two out of N strings $X_1, \ldots, X_N$ consisting of consecutive patterns B (i.e., long run of B's). To address this problem, we first build a trie on strings $X_1, \ldots, X_N$, and then ask queries with only one specified component, namely B. In other words, the search pattern is $\omega = (BBBBB\cdots)$. For example, for $B = 1$, and hence $\omega = (111111\cdots)$, we look for the right-most node in a trie built over N keys. Our Theorem 1(i) predicts that the average lengths of such long runs of 1's is $\log_{p_1} N^{-1} + O(1)$. If $B = 100$, then the longest runs are on the average equal to $\log N/\log(p_1 p_0^2)^{-1} + O(1)$, and so forth. An even more interesting problem is to study long runs of pattern B in a *single* string X. This is a well known problem in combinatorics and a number of papers were devoted to it (cf. [12], [15], [16]). Surprisingly enough, we can tackle it using the same approach as above. In this case, however, instead of constructing a trie we must build a suffix tree of X. A suffix tree of a string is a trie built from consecutive suffixes of the string (cf. [3], [4], [5]). Then, a partial match query searching for pattern B (i.e., $\omega = (BBB\cdots)$) will find the longest run of the pattern B. But it is known that suffix trees are statistically almost identical to tries (cf. [11] and [18]). In particular, the mean value of the longest run of the pattern B containing m zeros and $M - m$ ones is asymptotically equal to $\log N/\log(p_0^{-m} p_1^{m-M}) + O(1)$.

(ii) *A Composite Parameter in Tries.* If the match query contains at least one unspecified component, the asymptotics of the cost change from $O(\log N)$ to $O(N^{-\kappa_0})$ for $-1 < \kappa_0 < 0$ according to Theorem 1(ii). This is a simple consequence of a new nature of the parameter that one computes during the partial match. Note that when the query q consists only of unspecified components, then the cost of the partial match is exactly the number of internal nodes in the underlying trie, which is asymptotically $N(1+\xi(N))/h$ where h is the entropy of the alphabet and $\xi(N)$ a fluctuating function with small amplitude (cf. [25]). As discussed in (i) if the query q contains only specified components, then the cost corresponds to a *specific* path in the trie. Therefore, in the intermediate case the cost of a query defines a new parameter in a trie. It is composed of layers of internal nodes (when an unspecified component is executed) which are connected by pieces of a path in the trie. This parameter grows asymptotically like $O(N^{-\kappa_0})$. Moreover, Theorem 2 asserts that the cost is likely to be equal to the average value of the cost with high probability.

(iii) *Extensions.* Our results can be extended in many directions. For example, it is easy to see that they hold for V-ary symmetric tries, that is, when $\Sigma = \{1, 2, \cdots, V\}$. It may be interesting to extend our Theorem 2 to an asymmetric alphabet, and it is even more challenging to assume some dependency among symbols in the alphabet Σ. Finally, one may replace the regular tries by other digital structures such as Patricia tries and digital search trees. Some preliminary results in this direction are reported in [20]. Also, a rigorous analysis of suffix trees with partial match retrieval might be of some interest.

(iv) *Proof of Theorem 2(ii).* We just note that part (ii) of Theorem 2 follows directly from part (i) and Chebyshev's inequality.

3. ANALYSIS

In this section we prove Theorems 1 and 2.

3.1 The Average Cost in the Asymmetric Case

We start our analysis from equation (3) representing the recurrence equation for the average cost l_N^ω of the query ω. To solve this recurrence we use an old trick (cf. [21], [22] and [23]) with some modifications as in [13]. Let $L^\omega(z)$ be the exponential generating function of the cost, that is, $L^\omega(z) = \sum_{N=0}^{\infty} l_N^\omega z^N/N!$. The recurrence (3) can be easily transformed into the following functional equation

$$L^\omega(z) = \delta(0,\omega_1)L^{\omega'}(zp_0)e^{p_1 z} + \delta(1,\omega_1)L^{\omega'}(zp_1)e^{p_0 z} + b(z) \ ,$$

where $b(z) = e^z - 1 - z$. Introducing the transform $\tilde{L}^\omega(z) = e^{-z}L^\omega(z)$ we reduce the above to

$$\tilde{L}^\omega(z) = \delta(0,\omega_1)\tilde{L}^{\omega'}(zp_0) + \delta(1,\omega_1)\tilde{L}^{\omega'}(zp_1) + \tilde{b}(z) \ , \tag{14}$$

where $\tilde{b}(z) = e^{-z}b(z)$. Using the fact that $\omega = \omega^{(M)}$, and iterating (14) M times we finally obtain the following

$$\tilde{L}^\omega(z) = \sum_{\mathbf{b}\in\Theta_M} \delta(b_1,\omega_1)\cdots\delta(b_M,\omega_M)\tilde{L}^\omega(zp_{b_1}\cdots p_{b_M}) + \sum_{j=0}^{M-1}\sum_{\mathbf{b}\in\Theta_j} \delta(b_1,\omega_1)\cdots\delta(b_M,\omega_M)\tilde{b}(zp_{b_1}\cdots p_{b_M}) \ .$$

This functional equation is easy to solve. Comparing the coefficients of $\tilde{L}(z)$ at z^N we finally prove that

$$\tilde{l}_N^\omega = \frac{(-1)^N(N-1)\sum_{j=1}^{M-1}\sum_{\mathbf{b}\in\Theta_j}\prod_{i=1}^{j}\delta(b_i,\omega_i)p_{b_i}^N}{1 - \sum_{\mathbf{b}\in\Theta_M}\prod_{i=1}^{M}\delta(b_i,\omega_i)p_{b_i}^N} \ . \tag{15}$$

To recover the original average cost l_N^ω we note that $l_N^\omega = \sum_{k=1}^{N}\binom{N}{k}(-1)^{k+N}\tilde{l}_k^\omega$.

The asymptotics of (15) are easy to obtain since the general result of [27] (Mellin transform approach) or of [14] (Rice's formula) can be applied to the alternating sum for l_N^ω. In particular, using [27] we translate l_N^ω into the following Mellin-like integral

$$l_N^\omega \sim \frac{1}{2\pi i}\int_{-3/2-i\infty}^{-3/2+i\infty} \frac{\Gamma(z)N^{-z}(z+1)\sum_{j=1}^{M-1}\sum_{\mathbf{b}\in\Theta_j}\prod_{i=1}^{j}\delta(b_i,\omega_i)p_{b_i}^{-z}}{1-\sum_{\mathbf{b}\in\Theta_M}\prod_{i=1}^{M}\delta(b_i,\omega_i)p_{b_i}^{-z}}dz + O(1) .$$

This integral is easy to evaluate by Cauchy's residue theorem. Computing the residues of the integrand we finally obtain Theorem 1. Note that in the case $s = M$ there is double pole at $z = 0$ which produces the logarithm function in Theorem 1(i). In the case $0 < s < M$ the dominating pole is a single pole at $z = \kappa_0$. The fluctuating function $\xi(N)$ is a consequence of other poles at $\kappa_k = \kappa_0 + iy_k$ which are solutions of the equation (5) (the denominator of the above integrand). It is easy to prove that $-1 < \kappa_0 < 0$, as needed for Theorem 1(ii).

3.2 The Analysis of the Variance in the Symmetric Case

We first consider $\omega_1 = (*, S)$. For simplicity of notation we also write $\omega_1 = \omega$. Our task is to estimate the variance of the total cost C_N^ω. We use formula (8), and therefore we need to estimate the second factorial moment of the cost. In fact, before we launch an attack on it, we need more precise evaluation of the average cost l_N^ω. Both parameters will follow from the system of recurrences of the probability generating function given in (7).

Using (7) for $\omega = (*, S)$, we find that l_N^ω satisfies the following system of recurrences

$$\begin{aligned} l_N^\omega &= 1 + 2^{1-N}\sum_{k=0}^{N}\binom{N}{k} l_k^{\omega'} \\ l_N^{\omega'} &= 1 + 2^{-N}\sum_{k=0}^{N}\binom{N}{k} l_k^{\omega} \end{aligned}$$

Defining the exponential generating function $L_N^\omega(z)$ and its transform $\tilde{L}_N^\omega(z) = e^{-z}L^\omega(z)$, we obtain the following system of functional equations with $b(z) = 1-(1+z)e^{-z}$

$$\begin{aligned} \tilde{L}^\omega(z) &= 2\tilde{L}^{\omega'}\left(\frac{z}{2}\right) + b(z) \\ \tilde{L}^{\omega'}(z) &= \tilde{L}^{\omega}\left(\frac{z}{2}\right) + b(z) , \end{aligned}$$

which becomes after some algebra

$$\tilde{L}^\omega(z) = 2\tilde{L}^\omega\left(\frac{z}{4}\right) + b(z) + 2b\left(\frac{z}{2}\right) .$$

Solving it, one gets

$$\tilde{l}_N^\omega = \frac{(-1)^N(N-1)2^{1-N}}{1-2^{1-2N}} ,$$

and hence

$$l_N^\omega = \sum_{k=1}^{N} \binom{N}{k} (-1)^k f(k) \tag{16}$$

with analytical continuation of $f(k)$ as below

$$f(z) = \frac{(z-1)(1+2^{1-z})}{1-2^{1-2z}} \ . \tag{17}$$

The alternating sum can be treated in the same manner as in the previous section. For the reader's convenience we reproduce here a lemma that summarizes Rice's method of dealing with such alternating sums (cf. [14], [21], [22]).

Lemma. *Let C be a path surrounding the points $\{j, j+1, \dots, N\}$, and $f(z)$ be an analytical continuation of $f(k)$ inside C. Then*

$$\sum_{k=j}^{N} \binom{N}{k} (-1)^k f(k) = -\frac{1}{2\pi i} \int_C [N;z] f(z) dz \ , \tag{18}$$

where

$$[N;z] = \frac{\Gamma(-z)\Gamma(N+1)}{\Gamma(N+1-z)} = \frac{(-1)^{N-1} N!}{z(z-1)\cdots(z-N)} \ , \tag{19}$$

and $\Gamma(z)$ is the gamma function (cf. [1]*)* . ∎

Now it is easy to notice that the solution for l_N^ω fulfills the condition of the lemma. The function $f(z)$ in (17) has single poles at $z_k = \frac{1}{2} + \frac{1}{2}\chi_k$ where $\chi_k = 2\pi i k/L$ and $L = \log 2$. Computing the residues at z_k and additionally at $z = 0$ (coming from the function $[N;z]$), we finally obtain

$$l_N^\omega \sim \sqrt{N} \left(\sqrt{\pi} \frac{1+\sqrt{2}}{2L} - 3 + \tau^\omega \left(\log_2 \sqrt{N} \right) \right) \tag{20}$$

where the fluctuating function $\tau^\omega(x) = \sum_{k \neq 0} \tau_k^\omega e^{2k\pi i x}$ has the Fourier coefficients

$$\tau_k^\omega = \frac{1}{2L} \left(1 + \sqrt{2}(-1)^k \right) \Gamma\left(\frac{-1-\chi_k}{2} \right) \left(\frac{-1+\chi_k}{2} \right) \ . \tag{21}$$

Now we deal with the second factorial moment. Let $w_N^\omega = (H_N^\omega)''(1)$, and $W^\omega(z)$ be its exponential generating function. Traditionally, we also define $\tilde{W}^\omega(z) = W^\omega(z)e^{-z}$. From (7), after differentiating it twice, we obtain

$$\begin{aligned} \tilde{W}^\omega(z) &= 2\tilde{W}^{\omega'}\left(\frac{z}{2}\right) + 4\tilde{L}^{\omega'}\left(\frac{z}{2}\right) + 2\left(\tilde{L}^{\omega'}\left(\frac{z}{2}\right)\right)^2 \\ \tilde{W}^{\omega'}(z) &= \tilde{W}^{\omega}\left(\frac{z}{2}\right) + 2\tilde{L}^{\omega}\left(\frac{z}{2}\right) \ . \end{aligned}$$

For simplicity of presentation, we split $\tilde{W}^{\omega}(z)$, namely $\tilde{W}^{\omega}(z) = \tilde{R}^{\omega}(z) + \tilde{T}^{\omega}(z)$, where

$$\begin{aligned} \tilde{R}^{\omega}(z) &= 2\tilde{R}^{\omega'}\left(\frac{z}{2}\right) + 4\tilde{L}^{\omega'}\left(\frac{z}{2}\right) \\ \tilde{R}^{\omega'}(z) &= \tilde{R}^{\omega}\left(\frac{z}{2}\right) + 2\tilde{L}^{\omega}\left(\frac{z}{2}\right) \end{aligned} \tag{22}$$

and

$$\begin{aligned} \tilde{T}^{\omega}(z) &= 2\tilde{T}^{\omega'}\left(\frac{z}{2}\right) + 2\left(\tilde{L}^{\omega'}\left(\frac{z}{2}\right)\right)^2 \\ \tilde{T}^{\omega'}(z) &= \tilde{T}^{\omega}\left(\frac{z}{2}\right) . \end{aligned} \tag{23}$$

The solution of the first system of recurrences, namely (22), is easy. Note that it reduces to the following functional equation

$$\tilde{R}^{\omega}(z) = 2\tilde{R}^{\omega}\left(\frac{z}{4}\right) + 4\tilde{L}^{\omega'}\left(\frac{z}{2}\right) + 4\tilde{L}^{\omega}\left(\frac{z}{4}\right) ,$$

which we know how to solve. Repeating the arguments from [21] and [22] we immediately find that

$$\tilde{r}_N^{\omega} = \frac{(-1)^N (N-1) 2^{2-N}}{(1-2^{1-2N})^2}\left[1 + 2^{1-N} + 2^{1-2N}\right] . \tag{24}$$

Now we deal with (23), which is more intricate. To handle it we need a formula for $\tilde{M}^{\omega'}(z) = \left(\tilde{L}^{\omega'}(z)\right)^2$. But from

$$\tilde{L}^{\omega'}(z) = 2\tilde{L}^{\omega'}\left(\frac{z}{4}\right) + b(z) + b\left(\frac{z}{2}\right) ,$$

after taking it to the power two, we finally obtain

$$\tilde{M}^{\omega'}(z) = 4\tilde{M}^{\omega'}\left(\frac{z}{4}\right) + 4\left(b(z) + b\left(\frac{z}{2}\right)\right)\tilde{L}^{\omega'}\left(\frac{z}{4}\right) + \left(b(z) + b\left(\frac{z}{2}\right)\right)^2 .$$

This means that the coefficient $\tilde{m}_N^{\omega'}$ at $z^N/N!$ is $\tilde{m}_N^{\omega'} = \frac{1}{1-2^{2-2N}} \cdot (4A_N + B_N)$, where, after some tedious algebra, one finds

$$\begin{aligned} A_N &= \left[\frac{z^N}{N!}\right]\left(b(z) + b\left(\frac{z}{2}\right)\right)\tilde{L}^{\omega'}\left(\frac{z}{4}\right) \\ &= (-1)^N \sum_{l=1}^{N-1} \binom{N}{l} \frac{(l-1)(N-l-1)\left(1+2^{-l}\right)\left(1+2^{-N+l}\right)2^{-2(N-l)}}{1-2^{1-2(N-l)}} \\ &= (-1)^N \sum_{l=1}^{N-1} \binom{N}{l} \frac{(l-1)(N-l-1)\left(1+2^{-l}\right)\left(1+2^{-N+l}\right)2^{-2l}}{1-2^{1-2l}} , \end{aligned} \tag{25}$$

and

$$B_N = (-1)^N(N-1)\left(2^{N-2}(N-4) + \frac{1}{4}(N+12) + 2^{2-N} + \frac{1}{2}\left(\frac{3}{2}\right)^{N-2}(2N-9)\right) \ . \quad (26)$$

Putting everything together we find that $w_N^\omega = \sum_{k=2}^N \binom{N}{k}(-1)^k f(k)$ with the analytical continuation $f(z)$ of $f(k)$ as below

$$\begin{aligned} f(z) &= \frac{(z-1)2^{1-z}}{(1-2^{1-2z})(1-2^{2-2z})}\left[5 + 2^{3-z} + 2^{3-2z} + z2^{z-2} - 2^z + \frac{z}{4} + \left(z - \frac{9}{2}\right)\left(\frac{3}{2}\right)^{z-2}\right] \\ &+ \frac{2^{3-2z}}{(1-2^{1-2z})(1-2^{2-2z})}\sum_{l\geq 2}\binom{z}{l}\frac{(l-1)(z-l-1)\left(1+2^{-l}\right)\left(1+2^{z-l}\right)2^{-l}}{1-2^{1-2l}} \ . \quad (27) \end{aligned}$$

Now we can apply our Lemma to the alternating sum w_N^ω to obtain an asymptotic expansion. Note that there are two set of poles that contribute leading terms of the asymptotics. Namely, $z_{k,1} = 1 + \frac{1}{2}\chi_k$ and $z_{k,2} = \frac{1}{2} + \frac{1}{2}\chi_k$ where $k = 0, \pm 1, \cdots$. Easy computations reveal finally that

$$\begin{aligned} w_N^\omega &\sim \frac{N}{L}\left\{8 - \frac{7}{12} - 2\sum_{l\geq 2}(-1)^l \frac{\left(1+2^{-l}\right)\left(1+2^{1-l}\right)2^{-l}}{1-2^{1-2l}} + \tau_3(\log_2\sqrt{N})\right\} \\ &- 2\sqrt{\pi N}\left\{\frac{1}{L}\left[\frac{73}{2\sqrt{2}} + \frac{25}{16} - \frac{4}{3\sqrt{3}}\right]\right. \\ &+ \left.\frac{2}{L}\sum_{l\geq 2}\binom{1/2}{l}\frac{(l-1)\left(l+\frac{1}{2}\right)\left(1+2^{-l}\right)\left(1+2^{\frac{1}{2}-l}\right)2^{-l}}{1-2^{1-2l}} + \tau_4(\log_2\sqrt{N})\right\} \quad (28) \end{aligned}$$

with periodic fluctuations $\tau_3(x)$ and $\tau_4(x)$ of mean zero.

Now we are ready to put everything together. In particular, we note that, apart from the continuous fluctuations of mean zero,

$$\text{var } C_N^\omega = w_N^\omega + l_N^\omega - (l_N^\omega)^2 = BN + A_1\sqrt{N} + o(\sqrt{N}) \ , \quad (29)$$

where A_1 and B are constants. We prove now that $B = 0$. Note that

$$B = \frac{1}{L}\left\{8 - \frac{7}{12} - 2\sum_{l\geq 2}(-1)^l\frac{\left(1+2^{-l}\right)\left(1+2^{1-l}\right)2^{-l}}{1-2^{1-2l}}\right\} - \pi\frac{\left(1+\sqrt{2}\right)^2}{4L^2} + [(\tau^\omega)^2]_0 \ ,$$

where $[(\tau^\omega)^2]_0$ is the mean of the square of the periodic function $\tau^\omega(x)$ defined in (20). As in [21] and [22], we point out that this term significantly contributes to the cancellation of terms in B.

To find a transformation for $[(\tau^\omega)^2]_0$ we proceed as follows. Note that

$$[(\tau^\omega)^2]_0 = 2\sum_{k\geq 1}\tau_k^\omega\tau_{-k}^\omega = \frac{2}{4L^2}\sum_{k\geq 1}\left(3 + 2\sqrt{2}(-1)^k\right)\Gamma\left(\frac{1-\chi_k}{2}\right)\Gamma\left(\frac{1+\chi_k}{2}\right) \ .$$

Now we use the formula (cf. [1])

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z} \tag{30}$$

and after some algebra we obtain

$$\Gamma(\frac{1-\chi_k}{2})\Gamma(\frac{1+\chi_k}{2}) = \frac{\pi}{\sin(\pi/2 + ik\pi^2/L)} = \frac{\pi}{\cos(ik\pi^2/L)} = \frac{\pi}{\cosh(k\pi^2/L)} = 2\pi\frac{e^{-k\pi^2/L}}{1+e^{-2k\pi^2/L}}\,,$$

so that

$$[(\tau^\omega)^2]_0 = \frac{\pi}{L^2}\sum_{k\geq 1}\left(3+2\sqrt{2}(-1)^k\right)\frac{e^{-k\pi^2/L}}{1+e^{-2k\pi^2/L}}\,. \tag{31}$$

Let us define two new functions

$$F(x) = \sum_{k\geq 1}\frac{e^{-kx}}{1+e^{-2kx}}\,, \tag{32}$$

$$G(x) = \sum_{k\geq 1}\frac{(-1)^{k-1}e^{-kx}}{1+e^{-2kx}}\,. \tag{33}$$

Then, (31) in terms of $F(x)$ and $G(x)$ becomes

$$[(\tau^\omega)^2]_0 = \frac{3\pi}{L^2}F\left(\frac{\pi^2}{L}\right) - \frac{2\sqrt{2}\pi}{L}G\left(\frac{\pi^2}{L}\right)\,. \tag{34}$$

This form is not yet suitable for cancellation of terms at N in (29), and therefore we need to derive some series transformation for $F(x)$ and $G(x)$.

In order to find a transformation result for $F(x)$ we rewrite this function as

$$F(x) = \sum_{j\geq 1}(-1)^j\sum_{k\geq 1}e^{-k(2j+1)x} = \sum_{j\geq 0}\chi(j)\frac{1}{e^{jx}-1}$$

where

$$\chi(j) = \begin{cases} 0 & \text{for } j \text{ even} \\ 1 & \text{for } j \equiv 1 \bmod 4 \\ -1 & \text{for } j \equiv 3 \bmod 4 \end{cases}$$

In [8] on page 258 we find the following Entry 11 from *Ramanujan's Notebooks.*

For α, $\beta > 0$ such that $\alpha\beta = \pi$ and $y \in \Re$ with $|y| < \beta/2$ the following holds

$$\alpha\left\{\frac{1}{4}\sec(\alpha y) + \sum_{k=1}^{\infty}\chi(k)\frac{\cos(\alpha yk)}{e^{k\alpha^2}-1}\right\} = \beta\left\{\frac{1}{4} + \frac{1}{2}\sum_{k=1}^{\infty}\frac{\cosh(2\beta yk)}{\cosh(k\beta^2)}\right\}\,.$$

For $y = 0$ this reads

$$\alpha\left\{\frac{1}{4}+\sum_{k=1}^{\infty}\chi(k)\frac{1}{e^{k\alpha^2}-1}\right\} = \beta\left\{\frac{1}{4}+\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{\cosh(k\beta^2)}\right\} .$$

Replacing in the above $\cosh(x)$ by the exponential functions, expanding the geometric series and rearranging the sums we finally obtain

$$\frac{1}{2}\sum_{k\geq 1}\frac{1}{\cosh(k\beta^2)} = \sum_{j\geq 1}\chi(j)\frac{1}{e^{j\beta^2}-1} ,$$

so that, taking into account the constraint $\alpha\beta = \pi$, we find

$$\alpha\{\frac{1}{4}+F(\alpha^2)\} = \beta\{\frac{1}{4}+F(\beta^2)\} .$$

Now, we substitute in the above $\alpha = \sqrt{x}$ and $\beta = \pi/\sqrt{x}$. Then,

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{\pi}{x}F\left(\frac{\pi^2}{x}\right)$$

for $x > 0$, which is the desired transformation result. For $G(x)$ we observe that $G(x) = F(x) - 2F(2x)$, hence

$$G(x) = \frac{1}{4} + \frac{\pi}{x}F\left(\frac{\pi^2}{x}\right) - \frac{\pi}{x}F\left(\frac{\pi^2}{2x}\right) .$$

Applying the above to (34) we finally obtain

$$[(\tau^{\omega})^2]_0 = \frac{3}{4L} - \frac{\pi}{4L^2}\left(3+2\sqrt{2}\right) + \frac{3-2\sqrt{2}}{L}F(L) + \frac{2\sqrt{2}}{L}F\left(\frac{L}{2}\right) . \qquad (35)$$

In passing, we mention that the Appendix contains another derivation of the above series transformations. Our derivation can be used to produce some other identities, and it is of its own independent interest.

The last equation is used to prove the cancellation of all terms at N except an oscillating term of mean zero. The remaining function has to be identically equal to zero since otherwise the variance would be negative for infinitely many values of N, which is impossible. For more detailed arguments the reader is referred to our papers [21] and [22]. Thus, we have proved $B = 0$, and as a consequence of this var $C_N^{\omega} = O(\sqrt{N})$. The coefficient A_1 at $\sqrt{N}$ in (29) is given in Theorem 2(i) equation (10).

The second pattern $\omega_2 = (S, *)$ can be analyzed in a similar manner as above. We sketch only major differences. From Section 3.1 we may quickly conclude that the average cost in this case is

$$l_N^{\omega_2} \sim \sqrt{N}\left(\sqrt{\pi}\frac{1+\sqrt{2}}{2\sqrt{2}L} + \tau^{\omega_2}(\log_2\sqrt{N})\right) - 2$$

where the coefficients of $\tau_k^{\omega_2}$ become

$$\tau_k^{\omega_2} = \frac{1}{2\sqrt{2}L}\left(\sqrt{2} + (-1)^k\right)\Gamma\left(\frac{-1-\chi_k}{2}\right)\left(\frac{-1+\chi_k}{2}\right) ,$$

and simple analysis reveals that the zero terms belonging to the first pattern $\omega = \omega_1 = (*, S)$ and the second pattern satisfy the following relationship: $[(\tau^{\omega_2})^2]_0 = \frac{1}{2}[(\tau^{\omega_1})^2]_0$.

It is proved that the second factorial moment $w_N^{\omega_2}$ satisfies $w_N^{\omega_2} = \sum_{k=2}^{N}\binom{N}{k}(-1)^k f(k)$ where

$$\begin{aligned} f(z) &= \frac{(z-1)2^{1-2z}}{(1-2^{1-2z})(1-2^{2-2z})}\left[7 + z2^{z-2} + \frac{z}{4} + 2^{2-z} + \left(z - \frac{9}{2}\right)\left(\frac{3}{2}\right)^{z-2}\right] \\ &+ \frac{2^{3-3z}}{(1-2^{1-2z})(1-2^{2-2z})}\sum_{l\geq 2}\binom{z}{l}\frac{(l-1)(z-l-1)\left(1+2^{-l}\right)\left(1+2^{z-l}\right)2^{-l}}{1-2^{1-2l}} . \end{aligned}$$

The rest proceeds in exactly the same manner as before, and it leads to our second part of Theorem 2(i).

We finally mention, that it is not at all obvious how to extend this result to the case $M \geq 3$. The reader should note that the key argument in our proof of Theorem 2(i) is the transformation result (35). For $M \geq 3$ there is no comparable result in *Ramanujan's Notebooks*, and the alternative proof presented in the Appendix relies heavily on the duplication formula for the gamma function, which is not suitable for $M \geq 3$.

APPENDIX: Mellin Transform Approach to Ramanujan-Like Series Identities

In Section 3.2 we have used the following two series transformations

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{\pi}{x}F\left(\frac{\pi^2}{x}\right) , \tag{36}$$

$$G(x) = \frac{1}{4} + \frac{\pi}{x}F\left(\frac{\pi^2}{x}\right) - \frac{\pi}{x}F\left(\frac{\pi^2}{2x}\right) , \tag{37}$$

where $F(x)$ and $G(x)$ are defined in (32) and (33) respectively.

Most proofs of results like (36) use the theory of *modular functions* (cf. [8]). In this appendix we present an alternative approach by using *Mellin integrals*, a line of attack suggested to us by P. Flajolet.

To prove (36) we proceed as follows. Let

$$\beta(s) = \sum_{j\geq 0}(-1)^j\frac{1}{(2j+1)^s} .$$

Using definition (32) of $F(x)$, we have

$$F(x) = \sum_{k\geq 1} \frac{e^{-kx}}{1+e^{2kx}} = \sum_{j\geq 0}(-1)^j \sum_{k\geq 1} e^{-k(2j+1)x} ,$$

so that the Mellin transform $F^*(s) = \int_0^\infty F(x)x^{s-1}dx$ of $F(x)$ (cf. [10]) becomes

$$F^*(s) = \Gamma(s)\zeta(s)\beta(s) .$$

By the Mellin inversion formula this yields

$$F(x) = \frac{1}{2\pi i}\int_{3/2-i\infty}^{3/2+i\infty} \Gamma(s)\zeta(s)\beta(s)x^{-s}ds . \tag{38}$$

Now we take the two residues $s = 1$ and $s = 0$ out from the above integral to get

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \Gamma(s)x^{-s}\zeta(s)\beta(s)ds ,$$

where in the above we use the fact that $\beta(0) = 1/2$ and $\beta(1) = \pi/4$ (cf. [1]). Using the duplication formula for $\Gamma(s)$ (cf. [1]) we obtain

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \frac{1}{\sqrt{\pi}}2^{s-1}\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right)x^{-s}\zeta(s)\beta(s)ds .$$

We now use the functional equations for $\zeta(s)$ and $\beta(s)$, namely

$$\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{s-\frac{1}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s) \tag{39}$$

and

$$\beta(1-s)\Gamma\left(1-\frac{s}{2}\right) = 2^{2s-1}\pi^{-s+\frac{1}{2}}\Gamma\left(\frac{s+1}{2}\right)\beta(s) . \tag{40}$$

Identity (39) is Riemann's functional equation for $\zeta(s)$, and (40) is an immediate consequence of the functional equation for the Hurwitz's ζ-function $\zeta(s,a)$ (cf. [2]), and the fact that

$$\beta(s) = 4^{-s}[\zeta(s,\frac{1}{4}) - \zeta(s,\frac{3}{4})] .$$

Using (39) and (40) the inverse Mellin transform becomes

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2\pi i}\int_{-\frac{1}{2}-i\infty}^{-\frac{1}{2}+i\infty} \pi^{2s-1}\Gamma(1-s)x^{-s}\zeta(1-s)\beta(1-s)ds .$$

After the substitution $1-s=u$ we have

$$F(x)=\frac{\pi}{4x}-\frac{1}{4}+\frac{1}{2\pi i}\int_{\frac{3}{2}-i\infty}^{\frac{3}{2}+i\infty}\pi^{1-2u}\Gamma(u)x^{u-1}\zeta(u)\beta(u)du\ .$$

This proves (36). To prove (37) we just note that $G(x)=F(x)-2F(2x)$.

Using the above scheme we can prove *several other identities*. Below we just list some of them. Define

$$f(x) \;=\; \sum_{k\geq 1}\frac{e^{kx}}{\left(e^{kx}-1\right)^2}=\sum_{k\geq 1}\frac{k}{e^{kx}-1}, \tag{41}$$

$$g(x) \;=\; \sum_{k\geq 1}\frac{(-1)^{k-1}}{k\left(e^{kx}-1\right)}, \tag{42}$$

$$h(x) \;=\; \sum_{k\geq 1}\frac{1}{k\left(e^{2k\pi x}-1\right)}\ . \tag{43}$$

Then, the following identities hold:

1.
$$1+2\sum_{n\geq 1}e^{-\pi n^2 x}=\frac{1}{\sqrt{x}}+\frac{2}{\sqrt{x}}\sum_{n\geq 1}e^{-\pi n^2/x} \qquad \text{Jacobi's } \vartheta\text{–function,}$$

2.
$$f(x)=\frac{1}{x^2}\frac{\pi^2}{6}-\frac{1}{2x}+\frac{1}{24}-\frac{4\pi^2}{x^2}f\left(\frac{4\pi^2}{x}\right) \qquad \text{Ramanujan [8],}$$

3.
$$g(x)=\frac{\pi^2}{12x}-\frac{\log 2}{2}+\frac{x}{24}-g\left(\frac{2\pi^2}{x}\right) \qquad \text{Ramanujan [8],}$$

4.
$$h(x)=\frac{\pi}{12}\left(\frac{1}{x}-x\right)+\frac{\log x}{2}+h\left(\frac{1}{x}\right) \qquad \text{Dedekind and Ramanujan [8].}$$

The latter three identities play a prominent role in the analysis of the variance for some parameters of tries (cf. [19], [21], [22]).

Remark. It is fairly well known that the functional equations for the Riemann ζ-function and the ϑ-function can be derived from each other via the Mellin transform (see e.g., [6], [8]). Here we would like to point out the analogous ϑ-like identity that is related to the functional equation for $\beta(s)$. Let

$$H(x)=\sum_{j\geq 0}(-1)^j(2j+1)e^{-(2j+1)^2x}\ ,$$

then

$$H(x) = \frac{1}{8}\left(\frac{\pi}{x}\right)^{3/2} H\left(\frac{\pi^2}{16x}\right) \ .$$

ACKNOWLEDGMENT

We would like to thank Dr. Philippe Flajolet, INRIA, for his long standing support of our work, and several fruitful discussions on the subject of this paper.

References

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

[2] T. Apostol, *Introduction to Analytical Number Theory*, Springer-Verlag, New York 1976.

[3] A. Apostolico, The Myriad Virtues of Suffix Trees, *Combinatorial Algorithms on Words,* pp. 8596, Springer-Verlag, ASI F12 (1985).

[4] A. Apostolico and W. Szpankowski, Self-Alignments in Words and Their Applications, J. Algorithms, to appear.

[5] A.V. Aho, J.E. Hopcroft and J.D. Ullman, *The Design and Analysis of Computer Algorithms,* Addison-Wesley, Reading Mass. 1974.

[6] R. Bellman, *Analytic Number Theory – An Introduction*, Benjamin-Cummings, London 1980.

[7] J. L. Bentley, Multidimensional Binary Search Trees Used for Associative Searching, *Communication of the ACM*, 18, 509-517, 1975.

[8] B.C. Berndt, *Ramanujan's Notebooks. Part II*, Springer-Verlag, New York 1989.

[9] J. Borwein and P. Borwein, *Pi and the AGM*, John Wiley & Sons, New York 1987.

[10] B. Davis, *Integral Transforms and Their Applications*, Springer-Verlag, New York 1978.

[11] L. Devroye, W. Szpankowski and B. Rais, A Note on the Height on Suffix Trees, *SIAM J. Computing*, to appear.

[12] P. Erdös and P. Révész, On the Length of the Longest Head of Run, *Colloquia Math. Soc. J. Bolyai 16 Keszthely*, 219-228, 1975

[13] P. Flajolet and C. Puech, Partial Match Retrieval of Multidimensional Data, *J. of the ACM*, 33, 371-407, 1986.

[14] P. Flajolet and R. Sedgewick, Digital Search Trees Revisited, *SIAM J. Computing*, 15, 748-767, 1986.

[15] L. Gordon, M. Schilling and M. Waterman, An Extreme Value Theory for Long Head Runs, *Probab. Theory and Related Topics*, 72, 279-287, 1986

[16] L. Guibas and A.M. Odlyzko, A Long Repetitive Patterns in Random Sequences, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 38, 241-262, 1980

[17] P. Henrici, *Applied and Computational Complex Analysis*, John Wiley&Sons, New York 1977.

[18] P. Jacquet and W. Szpankowski, Autocorrelation on Words and Its Applications, INRIA Technical Report, No. 1106, 1989.

[19] P. Kirschenhofer and H. Prodinger, On Some Applications of Formulae of Ramanujan in the Analysis of Algorithms, Mathematika, 38, 1991, to appear.

[20] P. Kirschenhofer and H. Prodinger, Multidimensional Digital Searching – Alternative Data Structures, preprint, Technical University of Vienna, 1991.

[21] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the Variance of the External Path length in a Symmetric Digital Tries, *Discrete Applied Mathematics*, 25, 129-143, 1989.

[22] P. Kirschenhofer, H. Prodinger and W. Szpankowski, On the balance Property of Patricia Tries: External Path Length Viewpoint, *Theoretical Computer Science*, 68, 1-17, 1989.

[23] D.E. Knuth, *The Art of Computer Programming. Vol. 3*, Addison-Wesley, Reading MA 1973.

[24] H. Mahmoud, *Naturally Growing Trees*, John Wiley & Sons, 1991 (in press).

[25] M. Régnier and P. Jacquet, New Results on the Size of Tries, *IEEE Trans. Information Theory*, 35, 203-205 (1989).

[26] R. L. Rivest, Partial-Match Retrieval Algorithm, *SIAM J. Computing*, 5, 19-50, 1976.

[27] W. Szpankowski, The Evaluation of an Alternating Sum with Applications to the Analysis of Some Data Structures, *Information Processing Letters*, 28, 13-19, 1988.