

# Hypergeometrics and the cost structure of quadrees

Philippe Flajolet, G. Labelle, L. Laforest, Bruno Salvy

► **To cite this version:**

Philippe Flajolet, G. Labelle, L. Laforest, Bruno Salvy. Hypergeometrics and the cost structure of quadrees. [Research Report] RR-2249, INRIA. 1994. <inria-00074422>

**HAL Id: inria-00074422**

**<https://hal.inria.fr/inria-00074422>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# *Hypergeometrics and the Cost Structure of Quadrees*

Philippe FLAJOLET  
Gilbert LABELLE  
Louise LAFOREST  
Bruno SALVY

**N ° 2249**

Avril 1994

PROGRAMME 2



*Rapport  
de recherche*

**1994**

# Hypergeometrics and the Cost Structure of Quadrees

Philippe Flajolet, Gilbert Labelle, Louise Laforest, Bruno Salvy  
Rocquencourt and Montreal

**Abstract.** *Several characteristic parameters of randomly grown quadrees of any dimension are analysed. Additive parameters have expectations whose generating functions are expressible in terms of generalized hypergeometric functions. A complex asymptotic process based on singularity analysis and integral representations akin to Mellin transforms leads to explicit values for various structure constants related to path length, retrieval costs, and storage occupation.*

## Hypergéométriques et structure de coût des arbres quadrants

**Résumé.** Plusieurs caractéristiques fondamentales des arbres quadrants obtenus par croissance aléatoire sont analysées. Les paramètres additifs ont des espérances dont les séries génératrices s'expriment en terme de fonctions hypergéométriques. Un procédé d'analyse asymptotique complexe fondé sur l'analyse de singularité et sur des représentations intégrales apparentées à la transformation de Mellin conduit à des formes explicites pour diverses constantes de structure associées à la longueur de cheminement, au coût de recherche, ou à l'occupation mémoire.

# HYPERGEOMETRICS AND THE COST STRUCTURE OF QUADTREES

PHILIPPE FLAJOLET, GILBERT LABELLE, LOUISE LAFOREST, AND BRUNO SALVY

April 23, 1994

**ABSTRACT.** Several characteristic parameters of randomly grown quadtrees of any dimension are analysed. Additive parameters have expectations whose generating functions are expressible in terms of generalized hypergeometric functions. A complex asymptotic process based on singularity analysis and integral representations akin to Mellin transforms leads to explicit values for various structure constants related to path length, retrieval costs, and storage occupation.

## INTRODUCTION

Quadtrees, invented by Finkel and Bentley [7], constitute a convenient data structure for storing and retrieving data from some multidimensional data space (usually embedded in  $\mathbb{R}^d$ ) that extends the familiar binary search tree structure. For instance they allow for  $\mathcal{O}(\log n)$  retrieval time, they present a fair amount of flexibility with respect to partial match queries where only a subset of the coordinates are specified, and they may be used as an index to access multidimensional data on secondary storage. The major applications are discussed in [16, Chap. 3] and reviewed extensively in Samet's books [26, 27].

From an analyst's standpoint, quadtrees lead to many challenging problems. The expectations of characteristic parameters satisfy probabilistic divide-and-conquer recurrences that are appreciably more complicated than the quicksort and binary search tree recurrences to which they reduce when the dimension  $d$  of the data space is equal to 1.

Finkel and Bentley followed by Gonnet [15] conjectured, on the basis of extensive simulations, that the expected cost of a random search in a quadtree built from  $n$  random independent data of  $d$ -dimensional space grows like

$$(1) \quad \mathbf{E}\{C_n\} = \lambda_d \log n + \mu_d + o(1).$$

That conjecture, as regards the dominant term at least, was solved independently by Flajolet *et al.* [8] and Devroye-Laforest [4] who established that

$$\lambda_d = \frac{2}{d}.$$

In dimension  $d = 2$ , it was found by [4, 8] that the expectation admits an explicit form involving the harmonic number  $H_n$  from which there results that  $\mu_2$  is also known. However, for higher dimensions ( $d \geq 3$ ), only first order asymptotics have been obtained explicitly till now as the treatment of [4]

---

This work was partly supported by the ESPRIT Basic Research Action of the E. C. No. 7141 (ALCOM II) (P. Flajolet and B. Salvy) and by NSERC Grants OGP0105582 (L. Laforest) and OGP0107733 (G. Labelle).

had to rely on probabilistic methods while that of [8] had to be based on a somewhat qualitative asymptotic study of singularities of a differential system in the complex plane. Some of these basic results are surveyed in Mahmoud's informative book [25].

The results of [4, 8] have been extended in various ways, so that more probabilistic properties of random quadtrees are now known. This includes (i) statistics of node degrees for dimension 2, by Laforest and Labelle [22, 23]; (ii) storage occupation constants in dimension 2, by Hoshi and Flajolet [18]; (iii) the limit distribution of search costs for all dimensions by Flajolet and Lafforgue [9]. In addition, probabilistic estimates for the height of quadtrees have been developed by Devroye [3]. These results are briefly reviewed at the end of the paper (Section 7).

This paper proposes to show that the expectations of additive cost measures that are of interest in the analysis of quadtrees have explicit expressions. Thus, parameters hitherto only partly accessible via probabilistic methods or indirectly accessible through implicit generating function equations can be precisely quantified. From the methods developed in this work, an explicit form for the constant  $\mu_d$  of (1) is derived (Theorem 3); the probability distribution of the search cost  $C_n$  reduces to convolutions of Stirling numbers (Theorem 8); storage occupation constants for paged index trees of all dimensions can be computed (Theorems 6, 7); finally, the main constant intervening in the search of extreme points is determined (Theorem 9).

## 1. OUTLINE OF THE ANALYSIS

We first recall the principle of the quadtree data structure for multidimensional data access [7, 16, 26, 27]. In dimension  $d = 2$ , a sequence of points  $\mathcal{P} = (P_1, P_2, \dots, P_n)$ , where each  $P_j \in \mathbb{R}^2$ , is represented by a quaternary tree constructed inductively by the following rules.

- If  $n = 0$  ( $\mathcal{P} = \emptyset$ ), then the tree is empty.
- If  $n \geq 1$ , then the first point  $P_1$  is made the root of the tree. The four root subtrees are then made recursively from the four disjoint sublists of points

$$\mathcal{P}_{NW}, \mathcal{P}_{NE}, \mathcal{P}_{SW}, \mathcal{P}_{SE},$$

defined by restricting  $\mathcal{P} \setminus \{P_1\}$  to the four quadrants ( $NW, NE, SW, SE$ ) that are determined by the root node  $P_1$ .

The construction generalizes to any dimension  $d$  by considering the  $2^d$  regions (octants) determined by the root node; for data in  $\mathbb{R}^d$ , the quadtree is then a  $2^d$ -ary tree.

We analyze here randomly grown quadtrees built from  $n$  uniformly distributed random points of the unit hypercube  $[0, 1]^d$  of  $d$ -dimensional space. As is well known, this model is itself equivalent to a general model of data points with coordinates drawn independently from an arbitrary continuous distribution. Under this model, the probability that a tree of size  $n$  has a designated subtree (for instance the first root subtree) of size  $k$  is [8, 23]

$$(2) \quad \pi_{n,k} = \frac{1}{n} \sum_{\mathcal{L}} \frac{1}{i_1 \cdots i_{d-1}} \quad \mathcal{L} = \{(i_1, i_2, \dots, i_{d-1}) \mid k < i_1 \leq i_2 \leq \dots \leq i_{d-1} \leq n\}.$$

Several equivalent forms are available (see Appendix) involving either binomial coefficients or generalized harmonic numbers. The  $\pi_{n,k}$  fully characterize the random quadtree process.

After these preliminaries, we turn to a succinct description of the main thread of the paper.

*1. Recurrences and generating functions.* Consider an additive parameter of quadtrees like path length, or number of leaf nodes (see Section 2 for a definition). Its expectation  $f_n$  taken over

randomly grown trees of size  $n$  satisfies a probabilistic divide-and-conquer recurrence of the form

$$(3) \quad f_n = t_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k} f_k.$$

There,  $t_n$  represents a simple enough sequence, the ‘‘toll’’, which is  $n$  for path length or  $\delta_{n,1}$  for the number of leaves; the ‘‘splitting probabilities’’  $\pi_{n,k}$  are as given in (2).

When going to the realm of generating functions, the divide-and-conquer recurrence (3) translates into an integral equation (itself reducible to a linear differential equation) for the ordinary generating function of the expectations. With

$$(4) \quad f(z) = \sum_{n=0}^{\infty} f_n z^n,$$

one finds by the form (2) of the splitting probabilities

$$(5) \quad (z(1-z)D)^d [f(z) - t(z)] - 2^d f(z) = 0 \quad \text{with} \quad D = \frac{d}{dz},$$

where  $t(z)$  is the ordinary generating function of the toll sequence  $\{t_n\}$ . A direct analysis of these linear differential equations of order  $d$  is the approach underlying the treatment of [8, 9, 18].

2. *The Euler transform and hypergeometric forms.* A key step in this paper is to introduce on generating functions the Euler transformation

$$(6) \quad f^*(z) \equiv \mathcal{E} f(z) = \frac{1}{1-z} f\left(\frac{z}{z-1}\right)$$

which is well-known from the calculus of finite differences<sup>1</sup> or convergence acceleration processes. Under this transformation, the differential equation satisfied by the generating function of expectations assumes a particularly simple form,

$$(7) \quad (zD)^d [(1-z)(f^*(z) - t^*(z))] + 2^d z f^*(z) = 0.$$

The transformed equation is in turn equivalent to an inhomogeneous linear recurrence of order 1 for the coefficients  $f_n^* = [z^n] f^*(z)$

$$(8) \quad f_n^* = u_n + \left(1 - \frac{2^d}{n^d}\right) f_{n-1}^*,$$

for some known sequence  $u_n$  elementarily related to the tolls  $t_n$ .

For basic parameters of interest, the numbers  $t_n$  (hence also the  $u_n$ ) are simple, so that the coefficients  $f_n^*$  admit an explicit form involving the products

$$(9) \quad \left(1 - \frac{2^d}{3^d}\right) \left(1 - \frac{2^d}{4^d}\right) \cdots \left(1 - \frac{2^d}{n^d}\right).$$

Thus, the  $f_n^*$  are closely related to  ${}_dF_{d-1}$  hypergeometric functions. The  $f_n$  are recovered from the  $f_n^*$  by the Euler transformation again, as  $\mathcal{E}$  is involutive, which at the level of coefficients yields

$$(10) \quad f_n = \sum_k \binom{n}{k} (-1)^k f_k^*.$$

---

<sup>1</sup>The idea of introducing this transformation was first suggested empirically by numerical experiments with differences aimed at finding alternative representations for expectations of quadtree parameters.

From there explicit hypergeometric-like summations result for the expected values of search cost, number of leaves, and so on. These general algebraic manipulations are the subject of Section 2. A convenient notation for manipulating the special hypergeometric forms encountered is introduced in Section 3.

*3. Complex variables asymptotics.* The alternating sum (10) does not lend itself to elementary asymptotics as it involves exponentially large cancellations: the binomial coefficients of (10) come close to  $2^n$  while the  $f_n$  of interest are known *a priori* to be polynomially bounded.

We approach the problem by means of the singularity analysis method [10]: the behaviour of  $f_n$  as  $n \rightarrow \infty$  is dictated by the behaviour of  $f(z)$  as  $z \rightarrow 1$  from the left, which in turn relates via the Euler transformation of (6) to the behaviour of  $f^*(z)$  as  $z \rightarrow -\infty$ .

The first problem is thus to estimate asymptotically the alternating series

$$f^*(-t) = \sum_n f_n^*(-t)^n \quad \text{as } t \rightarrow \infty,$$

where, broadly speaking,  $f_n^*$  is of hypergeometric type because of the occurrence of the products (9). We take a route based on contour integration and the integral representation

$$(11) \quad f^*(-t) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} \varphi(s) t^s \frac{\pi ds}{\sin \pi s}.$$

There  $\varphi(s)$  is an ‘‘analytic extrapolation’’ of  $f_n^*$  in the sense that  $f_n^* = \varphi(n)$ . Such representations, akin to Mellin transforms, are also familiar from complex variables versions of Euler-Maclaurin summation *a la* Abel-Plana, as well as from the classical theory of ordinary and basic hypergeometric functions *a la* Mellin-Barnes [14, 28, 29].

Shifting the line of integration to the right in (11) provides for the asymptotic expansion of  $f^*(-t)$  as  $t \rightarrow \infty$  by the residue theorem. The expansion of  $f^*(-t)$  translates into an asymptotic expansion of  $f(z)$  as  $z \rightarrow 1$  (the Euler transformation exchanges 1 and  $\infty$ ).

Eventually, precise asymptotics for the coefficients  $f_n$  result by singularity analysis [10] (akin to Darboux’s method) that, under conditions of analytic continuation, justify the general implication

$$(12) \quad f(z) \sim \lambda(1-z)^{-\alpha} \left( \log \frac{1}{1-z} \right)^k \quad \Longrightarrow \quad [z^n] f(z) \sim \lambda \frac{n^{\alpha-1}}{\Gamma(\alpha)} (\log n)^k.$$

We are going to carry out this programme in detail in the case of some of the most important parameters of quadtrees. The simplest situation is that of path length which is discussed in Section 4. Slightly more sophisticated analytic extrapolation techniques are needed in the case of the paging constants of Section 5. Section 6 discusses further applications of the method to extreme points and the distribution of search costs.

## 2. BASIC RECURRENCES

A tree parameter  $\xi[a]$  is a function that maps trees ( $2^d$ -ary trees for quadtrees in dimension  $d$ ) to real numbers. Parameters studied here are *additive parameters* that are specified recursively by

$$(13) \quad \begin{cases} \xi[a] = t_{|a|} + \sum_{j=1}^{2^d} \xi[a_j] & \text{(the } a_j \text{ are the root subtrees of } a) \\ \xi[\emptyset] = t_0 & \text{(\emptyset is the empty tree),} \end{cases}$$

where  $\{t_n\}$  is a fixed number sequence. The size  $|a|$  of a tree  $a$  is as usual the number of internal nodes it contains which for quadtrees coincides with the number of data points that compose it.

Additive parameters are thus obtained by adding a “toll”  $t_n$  depending only on the size  $n$  of the tree and the values of the parameter on the root subtrees. For instance the number of leaves is obtained with  $t_n = \delta_{n,1}$  and internal path length with  $t_n = n$ .

The problem addressed here is to analyze the expectation  $f_n$  of such a parameter under the probabilistic model where a quadtree is randomly grown by  $n$  successive insertions of independently and uniformly distributed points of the unit hypercube  $[0,1]^d$ . This is the standard model for such analyses, see [4, 8, 23, 25]. Under the quadtree growth process the expectation  $f_n$  obey the recurrence that reflects the inductive definition (13),

$$(14) \quad f_n = t_n + 2^d \sum_{k=0}^{n-1} \pi_{n,k} f_k \quad \text{with} \quad \pi_{n,k} = \frac{1}{n} \sum_{k < i_1 \leq i_2 \leq \dots \leq i_{d-1} \leq n} \frac{1}{i_1 \cdots i_{d-1}}.$$

There, and throughout the paper,  $d$  denotes the dimension:  $d = 1$  for binary search trees and  $d = 2$  for standard quadtrees; the  $\pi_{n,k}$  are the fundamental splitting probabilities already quoted in (2).

Taking generating functions, formula (14) translates into a simple integral equation,

$$(15) \quad f(z) = t(z) + 2^d \mathbf{J}^{d-1} \mathbf{I} f(z),$$

where  $\mathbf{J}$  and  $\mathbf{I}$  are the following integral operators

$$\mathbf{I} f(z) = \int_0^z f(t) \frac{dt}{1-t}, \quad \mathbf{J} f(z) = \int_0^z f(t) \frac{dt}{t(1-t)}.$$

Equation (15) is readily translated into a linear differential equation of order  $d$ :

$$\mathbf{I}^{-1} \mathbf{J}^{1-d} [f(z) - t(z)] = 2^d f(z),$$

which is equivalent to the form given in (5).

The Euler transformation

$$f^*(z) \equiv \mathcal{E} f(z) = (1-Z)f(Z) \quad \text{with} \quad Z = \frac{z}{z-1}$$

is an involutive transformation of series,  $f = \mathcal{E}^2 f$ . When applied to the integral equations describing additive cost parameters of quadtrees, it appears to simplify in an essential way the analysis.

Here, we have by involutivity  $f(z) = (1-Z)f^*(Z)$ . Rewriting similarly  $t(z) = (1-Z)t^*(Z)$  transforms (15) into

$$(16) \quad (1-Z)f^*(Z) = (1-Z)t^*(Z) + 2^d \mathbf{J}^{d-1} \mathbf{I} [(1-Z)f^*(Z)].$$

The integral operators  $\mathbf{I}, \mathbf{J}$  simplify under the substitution  $z \mapsto Z(z)$  inside their function's argument:

$$(17) \quad \begin{aligned} \mathbf{I} [(1-Z(z))g(Z(z))] &= \int_0^z (1-Z(t))g(Z(t)) \frac{dt}{1-t} = \int_0^{Z(z)} -g(u) du \\ \mathbf{J} [g(Z(z))] &= \int_0^{Z(z)} g(u) \frac{du}{u}. \end{aligned}$$

Thus, with these transformation rules, and with  $Z$  taken as the independent variable, (16) becomes

$$(18) \quad (1-Z)f^*(Z) = (1-Z)t^*(Z) + 2^d \widehat{\mathbf{J}}^{d-1} \widehat{\mathbf{I}} f^*(Z),$$

where

$$\widehat{\mathbf{I}} g(Z) = - \int_0^Z g(u) du, \quad \widehat{\mathbf{J}} g(Z) = \int_0^Z g(u) \frac{du}{u}.$$



The form (18) is equivalent to

$$(19) \quad \widehat{\mathbf{I}}^{-1} \widehat{\mathbf{J}}^{1-d} [(1-Z)(f^*(Z) - t^*(Z))] = 2^d f^*(Z),$$

itself a rephrasing of (7).

The action of these operators on the Taylor expansion of  $f$  in the variable  $Z$  is particularly simple and taking coefficients of  $Z^n$  on both sides of (16) yields the *linear recurrence of order 1* that is basic to our treatment,

$$f_n^* - f_{n-1}^* = t_n^* - t_{n-1}^* - 2^d \frac{f_{n-1}^*}{n^d},$$

or equivalently (this justifies (8))

$$(20) \quad f_n^* = (t_n^* - t_{n-1}^*) + \left(1 - \frac{2^d}{n^d}\right) f_{n-1}^*.$$

An explicit form for the coefficients  $f_n^*$  is then available:  $f_0^* = f_0$  and  $f_1^* = f_0 - f_1$  follow from the definition of  $\mathcal{E}$ , while iterating (20) yields for  $n \geq 2$

$$(21) \quad f_n^* = \Omega_n \sum_{j=2}^n \frac{t_j^* - t_{j-1}^*}{\Omega_j},$$

where

$$(22) \quad \Omega_n = \prod_{j=3}^n \left(1 - \frac{2^d}{j^d}\right) \quad \text{and} \quad \Omega_2 = 1.$$

( $\Omega_2$  is defined in accordance with the usual convention of empty products.) The  $\Omega_n$  are ubiquitous in the expectations of additive cost measures of quadrees; motivated by the needs our analysis, we develop a special set of notations in Section 3 according to which we write  $\Omega_n$  as a pseudo-factorial:  $\Omega_n = [n]!$ .

Finally, the coefficients of the Euler transform obey the classical binomial convolution formula

$$(23) \quad f_n^* = \sum_{k=0}^n \binom{n}{k} (-1)^k f_k.$$

By involutivity a similar formula, already stated in (10), makes it possible to recover  $f_n$  from  $f_n^*$ . The combination of (10) and (21) then yields an explicit form for the  $f_n$ .

**Theorem 1.** *The expectation  $f_n$  of an additive parameter defined by the toll function  $t_n$  satisfies*

$$f_n = t_0 + n((2^d - 1)t_0 + t_1) + \sum_{k=2}^n \binom{n}{k} (-1)^k [k]! \left[ \sum_{j=2}^k \frac{t_j^* - t_{j-1}^*}{[j]!} \right],$$

where

$$[n]! = \prod_{j=3}^n \left(1 - \frac{2^d}{j^d}\right), \quad [2]! = 1, \quad \text{and} \quad t_n^* = \sum_{k=0}^n \binom{n}{k} (-1)^k t_k.$$

One also has the useful equality

$$t_n^* - t_{n-1}^* = [z^n] t\left(\frac{z}{z-1}\right).$$

The rest of this paper is devoted to algebraic and asymptotic studies of  $f_n$  for various sequences  $t_n$ .

3. THE CALCULUS OF “ $d$ -ANALOGUES”

In order to manipulate expressions occurring in the analysis of quadtrees, it proves convenient to introduce a special set of notations inspired by the classic framework of  $q$ -analogues in combinatorial theory (though there is no close relation to our convention). This simplifies appreciably computations since it avoids considering sums and products of the  $\Gamma$  and  $\psi$  functions taken over sets of roots of unity.

We fix an integral dimension  $d$  and define the  $d$ -analogue of a complex number  $s$  as

$$[s] = 1 - \frac{2^d}{s^d}.$$

From this, the  $d$ -analogue of the factorial is introduced, first for integer  $n$ , as

$$(24) \quad [n]! = [3] \cdot [4] \cdots [n] = \left(1 - \frac{2^d}{3^d}\right) \left(1 - \frac{2^d}{4^d}\right) \cdots \left(1 - \frac{2^d}{n^d}\right) \quad \text{with} \quad [2]! = 1.$$

Clearly, the  $d$ -factorial of integral arguments satisfies the recurrence:  $[n]! = [n] \cdot [n-1]!$ .

The  $d$ -factorial can next be extended to complex values as the  $d$ -analogue of the Gamma function

$$(25) \quad [s]! = \prod_{j=1}^{\infty} \frac{[j+2]}{[j+s]},$$

and this expression is also related to the classical Gamma function (see below). By construction, it satisfies the basic functional equation

$$(26) \quad [s]! = [s] \cdot [s-1]!$$

By design, the function  $[s]!$  coincides at all integer points with the number sequence  $[n]!$  which is central in the analysis of quadtrees, as evidenced by Theorem 1. We also have

$$(27) \quad [+ \infty]! = [3] \cdot [4] \cdot [5] \cdots = \prod_{m=3}^{\infty} \left(1 - \frac{2^d}{m^d}\right),$$

and  $[s]!$  has a simple pole at all integers  $\{1, 0, -1, \dots\}$  since  $[2] = 0$ . More generally, there are simple poles at all the roots of the equations  $(s+j)^d = 2^d$ , for  $j = 0, 1, 2$ , etc.

A rôle is also played in subsequent formulæ by the logarithmic derivative of the  $d$ -factorial, called the  $d$ -analogue of the psi function,

$$(28) \quad \begin{aligned} [\psi](s+1) &= \frac{d}{ds} \log[s]! = - \sum_{j=1}^{\infty} \frac{d}{ds} \log[j+s] \\ &= -d2^d \sum_{j=1}^{\infty} \frac{1}{(j+s)^{d+1}[j+s]} \\ &= -d2^d \sum_{m=1}^{\infty} \frac{1}{(m+s)((m+s)^d - 2^d)}. \end{aligned}$$

In particular, we shall need a  $d$ -analogue of Euler's constant written  $[\gamma]$  (with a slight abuse of notation)

$$(29) \quad \begin{aligned} [\gamma] &= -[\psi](3) = d2^d \sum_{j=3}^{\infty} \frac{1}{j^{d+1}[j]} \\ &= d2^d \sum_{m=3}^{\infty} \frac{1}{m(m^d - 2^d)}, \end{aligned}$$

as well as the  $d$ -harmonic numbers:

$$(30) \quad [H]_n = d2^d \left[ \frac{1}{3^{d+1}[3]} + \frac{1}{4^{d+1}[4]} + \cdots + \frac{1}{n^{d+1}[n]} \right] = [\psi](n+1) + [\gamma], \quad [H]_2 = 0.$$

This formula serves to express the values of  $[\psi]$  at integer points (see Theorem 6).

The quantities  $[s]!$  and  $[\psi](s)$  are clearly related to the classical Gamma and psi functions. From the product formula for the Gamma function [29], one has

$$(31) \quad \frac{[s]!}{[+\infty]!} = \prod_{\omega^d=2^d} \frac{\Gamma(s+1-\omega)}{\Gamma(s+1)}, \quad [\psi](s) = -d\psi(s) + \sum_{\omega^d=2^d} \psi(s-\omega),$$

since  $\psi(s)$  is defined as the logarithmic derivative of  $\Gamma(s)$ . This entails in turn

$$(32) \quad \frac{1}{[+\infty]!} = \frac{1}{2^d} \prod_{\omega^d=2^d} \Gamma(3-\omega), \quad [\gamma] = d\left(\frac{3}{2} - \gamma\right) - \sum_{\omega^d=2^d} \psi(3-\omega).$$

Thus the quantities appearing in the statements of our theorems can be expressed directly as infinite sums and products (25,27,28,29) as well as evaluated numerically by a reduction to common special functions (31,32).

When  $d = 1$ , one has

$$\frac{[s]!}{[+\infty]!} = \frac{2}{s(s-1)},$$

from which the analysis of binary search trees is recovered as a particular case of our general theorems. In the case of  $d = 2$ , the  $d$ -analogue again degenerates into a rational function since  $\omega^2 = 2^2$  implies  $\omega = \pm 2$ , and one finds

$$\frac{[s]!}{[+\infty]!} = \frac{(s+1)(s+2)}{s(s-1)}.$$

In the perspective of this paper, this degeneracy ‘‘explains’’ that the analysis of standard quadtrees ( $d = 2$ ) is easier and elementary expressions are encountered as exemplified by [4, 8, 9, 18, 23].

#### 4. PATH LENGTH

Path length is defined as the additive parameter on trees corresponding to the toll function  $t_n = n$ , or equivalently as the sum of the depths of all internal nodes in the tree, with the depth of a node  $\nu$  being counted as the number of nodes along the branch connecting  $\nu$  to the root of the tree. (For instance, the root itself is at depth 1, and the path length of a tree of size 1 equals 1.)

Let  $f_n$  be the expectation of path length in a quadtree of size  $n$ . The cost of a random successful search in such a tree, as measured by the number of internal nodes traversed, is  $f_n/n$ .

We thus take  $t_n = n$  in (14). From the definition of  $\mathcal{E}$ , the derived sequence  $t_j^* - t_{j-1}^*$  is therefore  $\delta_{j,2} - \delta_{j,1}$  so that (21) reduces to

$$(33) \quad f_0^* = 0, \quad f_1^* = -1, \quad f_n^* = [n]! \quad (n \geq 2).$$

The generalized hypergeometric series  ${}_pF_q$  is classically defined [6, Vol. 1, Chap. V] by

$${}_pF_q \left( \begin{matrix} a_1, a_2, \dots, a_p \\ b_1, b_2, \dots, b_q \end{matrix} \middle| t \right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{t^n}{n!},$$

where  $(a)_\nu = \Gamma(a + \nu)/\Gamma(a)$  is the rising factorial. The form of (33) then corresponds to an explicit hypergeometric expression for the generating function  $f^*(t)$ :

$$(34) \quad f^*(t) = -t + t^2 {}_{d+1}F_d \left( \begin{matrix} 3 - \omega_1, \dots, 3 - \omega_d, 1 \\ 3, \dots, 3 \end{matrix} \middle| t \right),$$

where the  $\omega_j$  are the  $d$ -th roots of  $2^d$ . Application of the Euler transformation to (33,34) leads to the following theorem.

**Theorem 2.** *The expected internal path length of a quadtree of size  $n$  in dimension  $d$  is*

$$f_n = n + \sum_{k=2}^n \binom{n}{k} (-1)^k [k]!.$$

*The generating function of the sequence  $f_n$  is of hypergeometric type:*

$$f(z) = \frac{z}{(1-z)^2} + \frac{z^2}{(1-z)^3} {}_{d+1}F_d \left( \begin{matrix} 3 - \omega_1, \dots, 3 - \omega_d, 1 \\ 3, \dots, 3 \end{matrix} \middle| \frac{z}{z-1} \right) \quad \text{with} \quad \omega_j = 2e^{2ij\pi/d}.$$

Singularities of hypergeometric functions of type  ${}_{d+1}F_d$  are isolated and of algebraic-logarithmic type, as results from the differential equations that they satisfy. We can thus *a priori* use singularity analysis in order to estimate the asymptotic behaviour of  $f_n$  as  $n$  tends to infinity. This implies studying the behaviour of  $f(z)$  in the neighbourhood of 1 which in turn is governed by the behaviour of  $f^*(t)$  at  $-\infty$ .

Our goal is therefore to obtain a suitable analytic continuation for the hypergeometric series (34). While explicit formulae in terms of Meijer's  $G$  function are available by Slater's theorem [6, 28], we describe here a more direct method of attack which is applicable to other parameters of quadtrees for which no such simple hypergeometric form is available.

We first follow an idea originally developed by Mellin and Lindelöf that provides a complex integral representation for analytic functions whose coefficients admit analytic expressions.

**Lemma 1.** *Assume that an analytic function  $\varphi(s)$  satisfies the following conditions:*

- (i)  $\varphi(s)$  is analytic in the right half-plane  $\Re s \geq n_0 - \frac{1}{2}$ , for some integer  $n_0$ ;
- (ii) for some  $\theta < \pi$ , and some  $K > 0$ , inside the closed half-plane  $\Re s \geq n_0 - \frac{1}{2}$ ,  $\varphi(s)$  satisfies the bound

$$(35) \quad |\varphi(s)| < K e^{\theta|s|}.$$

*Then, the function  $F$  with Taylor coefficients  $(-1)^n \varphi(n)$  admits the integral representation:*

$$(36) \quad F(t) \equiv \sum_{n=n_0}^{\infty} \varphi(n) (-t)^n = -\frac{1}{2i\pi} \int_{n_0-1/2-i\infty}^{n_0-1/2+i\infty} \varphi(s) t^s \frac{\pi}{\sin \pi s} ds,$$

valid inside the sector  $|\arg(t)| \leq \pi - \theta'$ , for any  $\theta' > \theta$  and for  $|t| > T$  with  $T$  some fixed positive number.

**Proof.** (Sketch, see [24, p. 108–119] for details.) The validity of the expansion is proved by considering the integral of (36) taken along the contour formed with the line  $\Re(s) = n_0 - \frac{1}{2}$  and a large semi-circle extending to the right of  $\Re(s) = n_0 - \frac{1}{2}$  and centered at  $n_0 - \frac{1}{2}$ . The integral along that composite contour is then estimated by residues, with the integrand having a simple pole at  $s = n \in \{n_0, n_0 + 1, \dots\}$  with residue  $\varphi(n)(-t)^n$ .

The decay condition on  $\varphi(s)$  ensures convergence of both the integral and the sum in the stated region for  $t$ . Thus the integral equals the series.  $\square$

The integral representation of (36) is in fact an inverse Mellin integral and, as such, it usually lends itself to asymptotic analysis of  $F(t)$ , as  $t \rightarrow \infty$ , by shifting the line of integration to the left. Only a little more is needed to ensure the validity of the method.

**Lemma 2.** *Assume that  $\varphi(s)$  satisfies the conditions of Lemma 1, and in addition:*

- (i)  $\varphi(s)$  is meromorphic for  $\Re(s) \geq a$  for some  $a < 0$ ,
- (ii)  $\varphi(s)$  is analytic for  $\Re(s) = a$ .
- (iii)  $\varphi(s)$  is analytic in the domain defined by  $\Re(s) \geq a$  and  $|\Im(s)| \geq M$  for some  $M > 0$ .

Assume further that condition (35) holds for the extended analyticity domain of  $\varphi(s)$ . Then, the following expansion of  $F(t)$  holds:

$$F(t) \equiv \sum_{n=n_0}^{\infty} \varphi(n)(-t)^n = - \sum_{a < \Re(s) < n_0} \text{Res} \left( \varphi(s) t^s \frac{\pi}{\sin \pi s} \right) + \mathcal{O}(t^a) \text{ as } t \rightarrow \infty, \quad |\arg(t)| < \pi - \theta',$$

where the sum is taken over all poles in the strip  $a < \Re(s) < n_0$ .

**Proof.** The statement derives directly from a residue computation of the integral of (36) taken along a large rectangle extending towards  $\pm i\infty$  with vertical sides  $\Re(s) = a$  and  $\Re(s) = n_0 - \frac{1}{2}$ .  $\square$

Each residue contributes a term in the asymptotic expansion of  $F(t)$  towards  $\infty$ , with a pole of order  $k$  for  $\varphi(s)/\sin(\pi s)$  at  $s = \sigma$  producing a term of the form  $t^\sigma (\log t)^{k-1}$ . An intuitive statement of the chain formed with Lemmas 1 and 2 is then as follows:

*For an analytic function  $F(t)$  with coefficients that admit a suitable analytic form  $\varphi(s)$ , the asymptotic behaviour of  $F(t)$  at infinity derives from the singularities of the coefficient function  $\varphi(s)$ .*

This technique became classical a century ago. It is to be found in the books of Lindelöf [24] and Ford [13], as well as in Hardy's description [17, Chap. XI] of "Ramanujan's heuristic" for definite integration. For instance, eventually, the function

$$\log(1+t) = \sum_{n=1}^{\infty} (-1)^n \frac{t^n}{n}$$

is logarithmic as  $t \rightarrow \infty$  "because" its coefficient function  $\varphi(s) = \frac{1}{s}$  admits a simple pole at  $s = 0$ . (!) Related representations are of use in the study of classical or basic hypergeometric functions, see [5, Chap. II], [14, Chap. 4], [19, Sec. 3.5].

For path length, the function  $F(t) = t + f^*(-t)$  satisfies the conditions of the lemmas as its coefficients that involve  $[k]!$  are special cases of the  $d$ -analogue of the Gamma function  $[s]!$  introduced in (26). This provides an asymptotic expansion that translates directly into a matching expansion

for  $f(z)$  via the Euler transformation, and this in turn provides the sought asymptotic form of the coefficients  $f_n = [z^n]f(z)$  by singularity analysis.

Since here  $\varphi(s) = [s]!$ , we first take  $n_0 = 2$  and  $a = 1 - \epsilon$  where  $\epsilon$  is a small positive real number chosen in such a way that the equation  $s^d - 2^d = 0$  has only the root  $s = 2$  in  $\Re(s) > a$ . We have  $[s]! = \mathcal{O}(1)$  in  $\Re(s) > a$  and  $|\Im(s)| > 1$  as results from the expression of the  $d$ -analogue of the Gamma function in terms of the classical Gamma function (31), and from known growth properties of the Gamma function [29]. Thus, Lemmas 1 and 2 apply with  $\theta = 0$ .

For such an  $a = 1 - \epsilon$ , the only singularity encountered is a double pole at  $s = 1$  which induces a behaviour of the form  $t \log t$ : the residue is computed elementarily by first rewriting  $[s]! = [s + 1]!/[s + 1]$ , and using

$$\begin{aligned} [s + 1]! &= 1 - [\gamma](s - 1) + \mathcal{O}((s - 1)^2), & \frac{1}{[s + 1]} &= \frac{2}{d} \frac{1}{s - 1} + \frac{d + 1}{2d} + \mathcal{O}(s - 1), \\ t^s &= t + t \log t (s - 1) + \mathcal{O}((s - 1)^2), & \frac{\pi}{\sin \pi s} &= \frac{-1}{s - 1} + \mathcal{O}(s - 1), \end{aligned}$$

so that

$$-\text{Res} \left( [s + 1]! \frac{1}{[s + 1]} t^s \frac{\pi}{\sin \pi s} \right)_{s=1} = \frac{2}{d} t \log t + \frac{t}{2d} (d + 1 - 4[\gamma]).$$

The error term of  $\mathcal{O}(t^{1-\epsilon})$  can be improved. When shifting the contour further to the right, the next singularity which is encountered is either a double pole at the origin or two simple poles with real part  $1 - 2 \cos(2\pi/d)$ , so that adding these contributions to the residue at 1 we get

$$F(t) \equiv t + f^*(-t) = \frac{2}{d} t \log t + \frac{t}{2d} (d + 1 - 4[\gamma]) + \mathcal{O}(\log t + t^{1-2 \cos(2\pi/d)}), \quad t \rightarrow \infty.$$

Applying the Euler transform again, we get the behaviour of  $f$  at 1:

$$f(z) \sim \left[ \frac{2}{d} \log \frac{1}{1-z} + c \right] \frac{1}{(1-z)^2} \quad \text{with} \quad c = \frac{1}{2d} (3d + 1 - 4[\gamma]).$$

The basic method of singularity analysis [10] finally permits a termwise transfer from asymptotics of the generating function  $f(z)$  to the asymptotics of its coefficients,  $f_n$ . The translation is justified by the general theorem of singularity analysis [10], given that the asymptotic expansions of  $F(t)$  and  $f(z)$  hold in extended regions of the complex plane. We have

$$f(z) \sim \frac{2}{d} z \frac{d}{dz} \left[ \frac{1}{1-z} \log \frac{1}{1-z} \right] + \left( c - \frac{2}{d} \right) \frac{1}{(1-z)^2},$$

so that

$$[z^n]f(z) \sim \frac{2}{d} n H_n + \left( c - \frac{2}{d} \right) n.$$

**Theorem 3.** *The expected internal path length of a quadtree of size  $n$  in dimension  $d$  is asymptotically*

$$f_n = \frac{2}{d} n \log n + \mu_d n + \mathcal{O}(\log n + n^{-1+2 \cos(2\pi/d)}),$$

where  $\mu_d$  depends only on the dimension:

$$\mu_d = \frac{1}{2d} (3d - 3 + 4\gamma - 4[\gamma]) = \frac{3d - 3 + 4\gamma}{2d} - 2^{d+1} \sum_{m=3}^{\infty} \frac{1}{m(m^d - 2^d)}.$$

Theorem 3 is an improvement over the results of [4, 8] by which only dominant asymptotics could be derived explicitly. By shifting further the contour of integration in (36), one would even obtain in principle a full asymptotic expansion.

For binary search trees ( $d = 1$ ), we have

$$[\gamma] = \sum_{m=3}^{\infty} \frac{2}{m(m-2)} = \frac{3}{4},$$

which implies the known result [20, p. 427]:

$$f_n = 2n \log n + n(\gamma - 3) + \mathcal{O}(\log n).$$

For standard quadtrees ( $d = 2$ ), we find

$$[\gamma] = 8 \sum_{m=3}^{\infty} \frac{1}{m(m^2-4)} = \frac{11}{12}$$

and recover the asymptotic expansion of the exact form  $f_n = (n + 1/3)H_n - (n + 1)/6$  already obtained in [4, 8]:

$$f_n = n \log n + \left(\gamma - \frac{1}{6}\right)n + \mathcal{O}(\log n).$$

For  $d = 2, 3, 4, 5, 6$ , the numerical form of the asymptotic approximation of  $f_n/n$  is then:

$$\begin{aligned} d = 1 & : 2 \log(n) - 2.42278\ 43350\ 98467 \\ d = 2 & : \log(n) + 0.41054\ 89982\ 34866 \\ d = 3 & : \frac{2}{3} \log(n) + 0.97303\ 71351\ 30062 \\ d = 4 & : \frac{1}{2} \log(n) + 1.19708\ 16793\ 92943 \\ d = 5 & : \frac{2}{5} \log(n) + 1.30706\ 24971\ 52704 \\ d = 6 & : \frac{1}{3} \log(n) + 1.36793\ 35007\ 97805 \end{aligned}$$

These figures are in good agreement with the empirical formulæ inferred from simulations or numerical computations in [15, p.114] ( $\mu_3 \approx 0.997$ ) and [16, p. 145] ( $\mu_3 \approx 0.973036$ ).

## 5. PAGED QUADTREES

A quadtree may be used as an *index* for accessing data of  $d$ -dimensional space. In that case, all maximal subtrees of size less than a fixed threshold  $b$  (the page or “bucket” capacity) are stored separately into pages. The internal nodes of the original quadtree that do not belong to pages then constitute the internal nodes of the index.

The index tree corresponding to a quadtree  $a$  is thus a  $2^d$ -ary tree with a number of internal nodes  $\chi$  defined in terms of  $a$  itself by the recursion:

$$\begin{cases} \chi[a] = 0 & \text{if } |a| \leq b \\ \chi[a] = 1 + \sum_i \chi[a_i] & \text{if } |a| > b. \end{cases}$$

By the conservation law for  $2^d$ -ary tree, the number of pages is linearly related to  $\chi$ , being

$$1 + (2^d - 1)\chi[a].$$

Thus the analysis of page occupancy (the mean number of pages) in paged quadtrees reduces to the study of a particular additive parameter on quadtrees defined by the toll sequence

$$t_0 = t_1 = \cdots = t_b = 0, \quad t_{b+1} = t_{b+2} = \cdots = 1.$$

So far this parameter has only been analysed in dimension  $d = 2$  by Hoshi and Flajolet [18] to which we also refer for a more thorough algorithmic discussion.

In the particular case when  $b = 1$ , the parameter  $\chi[a]$  counts the number of nonleaf nodes of tree  $a$ . Then  $\lambda[a] = |a| - \chi[a]$  is the number of leaves of tree  $a$ . The parameter  $\lambda$  can also be analysed directly by means of the recurrence

$$\begin{cases} \lambda[\emptyset] = 0 \\ \lambda[a] = 1 & \text{if } |a| = 1 \\ \lambda[a] = \sum_i \lambda[a_i] & \text{if } |a| > 1. \end{cases}$$

We shall carry out the analysis of the number of leaves  $\lambda$  first as computations reveal the spirit of the method while avoiding the unnecessary complications of the general case.

**Leaf statistics.** Let  $f_n$  denote the expected number of leaves in a randomly grown quadtree of size  $n$ . The basic recurrence is the quadtree recurrence with toll function  $t_n = \delta_{n,1}$ , and initial conditions  $f_0 = 0$ ,  $f_1 = 1$ . Thus, Theorem 1 applies with the Euler transform of the toll function satisfying

$$t_n^* - t_{n-1}^* = [z^n] \frac{z}{z-1} = -1.$$

**Theorem 4.** *The expected number of leaves of a randomly grown quadtree of size  $n$  is given by*

$$f_n = n - \sum_{k=2}^n \binom{n}{k} (-1)^k \left[ \sum_{j=2}^k \frac{[k]!}{[j]!} \right].$$

Laforest [23] first derived an equivalent explicit form when  $d = 2$ :

$$f_n = 8H_n^{(2)}(3n+1) + 11 - 39n - \frac{4}{n},$$

with  $H_n^{(2)} = \sum_{j=1}^n 1/j^2$ .

The asymptotic evaluation according to the methods of the previous section requires an analytic extrapolation of the sequence

$$(37) \quad \nu_k = \sum_{j=2}^k \frac{1}{[j]!},$$

that is to say an analytic function  $\xi(s)$  satisfying  $\xi(k) = \nu_k$  for  $k$  any integer  $\geq 2$ .

The basic technique here, which is familiar from the theory of infinite sums (or products), is that of “convergence terms” (or “convergence factors”) already illustrated by the construction of the  $d$ -Gamma function. Its spirit is summarized by the following lemma.



**Lemma 3.** *Let  $\{\mu_j\}_{j=n_0}^\infty$  be a number sequence with analytic extrapolation  $\omega(s)$ , so that  $\omega(j) = \mu_j$  for  $j \geq n_0$ . Assume that  $\omega(s)$  admits an asymptotic expansion*

$$\omega(s) = \alpha(s) + o(1) \quad \text{as } s \rightarrow +\infty$$

for some polynomial  $\alpha(s)$ . Consider the summatory sequence  $\nu_k = \sum_{j=n_0}^k \mu_j$ .

Then, provided the sum below converges in some region of the complex  $s$ -plane, an analytic extrapolation  $\xi(s)$  of the sequence  $\{\nu_k\}_{k=n_0}^\infty$  is given by

$$\xi(s) = A(s) - A(n_0 - 1) + \sum_{j=n_0}^{\infty} [\omega(j) - \alpha(j) - \omega(s + j + 1 - n_0) + \alpha(s + j + 1 - n_0)],$$

where  $A(s)$  is a polynomial satisfying  $A(s) - A(s - 1) = \alpha(s)$ .

The proof is a simple verification as the finite sums “telescope”. The difficulty comes from the fact that it is no longer assumed that  $\mu_n \rightarrow 0$ . For instance, if  $\mu_n$  tends to a constant  $\mu$ , one should take  $\alpha(s) = \mu$  and  $A(s) = \mu s$ .

For leaves, by (37), we thus need to extrapolate a sum  $\nu_k$  whose general term,  $\mu_j = 1/[j]!$  tends to  $1/[+\infty]!$  as  $j \rightarrow +\infty$ . Thus, by Lemma 3, an analytic extrapolation of  $\nu_k$  is provided by

$$\xi(s) = \frac{(s-1)}{[+\infty]!} + \sum_{j=2}^{\infty} \left( \frac{1}{[j]!} - \frac{1}{[s+j-1]!} \right).$$

The rightmost pole encountered is at  $s = 1$ . There,  $\xi(1) = 0$  as the sum telescopes and

$$\xi(s) \sim \xi'(1)(s-1) \quad \text{with} \quad \xi'(1) = \frac{1}{[+\infty]!} + \sum_{j=2}^{\infty} \frac{[\psi](j+1)}{[j]!}.$$

The function extrapolating  $[k]! \cdot \nu_k$  is then  $[s]! \cdot \xi(s)$ .

The rightmost pole of interest (we have  $n_0 = 2$  in the notations of Lemma 3) is thus at  $s = 1$ , near which we have

$$\frac{\pi}{\sin \pi s} \sim -\frac{1}{s-1}, \quad [s]! \sim \frac{2}{d} \frac{1}{s-1}, \quad \xi(s) \sim \xi'(1)(s-1).$$

Thus, one finds for the mean number of nonleaf nodes

$$\begin{aligned} f^*(-t) &\sim \frac{2}{d} \xi'(1)t && (t \rightarrow +\infty) \\ f(z) &\sim \frac{2}{d} \xi'(1) \frac{1}{(1-z)^2} && (z \rightarrow 1) \\ f_n &\sim \frac{2}{d} \xi'(1)n && (n \rightarrow \infty), \end{aligned}$$

by the residue computation of Lemma 2, by application of the Euler transform, and finally by singularity analysis.

**Theorem 5.** *The mean number of leaves of a random quadtree of size  $n$  is asymptotic to*

$$C \cdot n \quad \text{where} \quad C = 1 - \frac{2}{d} \left( \frac{1}{[+\infty]!} + \sum_{j=2}^{\infty} \frac{[\psi](j+1)}{[j]!} \right),$$

where  $[s]!$  and  $[\psi](s)$  are defined in (25-32).

In dimension 1, the constant is known to be  $C = 1/3$ . In dimension 2, the proportion of leaf nodes simplifies to  $C = 4\pi^2 - 39 \doteq 0.47841$ , a result of [18, 23]. In higher dimensions, the constants are expressed by infinite sums that do not seem to reduce to known quantities.

**Paging.** The analysis of paging generalizes the expressions found in Theorems 4 and 5. It follows once more the paradigm of our earlier analyses that we have encapsulated in Figure 1. Given this schema, we can limit ourselves to stating only the main steps in the proofs of Theorems 6–9.

**Theorem 6.** (i). *The expected number of pages in a paged quadtree of size  $n$  and page size  $b$  in dimension  $d$  is*

$$f_n^{[b]} = 1 + (2^d - 1) \sum_{k=b}^n \binom{n}{k} (-1)^{k+b+1} \left[ \sum_{j=b+1}^k \binom{j-2}{b-1} \frac{[k]!}{[j]!} \right].$$

(ii). *Furthermore, there exists a computable constant  $C^{[b]}$  such that*

$$\lim_{n \rightarrow +\infty} \frac{f_n^{[b]}}{n} = C^{[b]}.$$

**Proof.** (i). We use the toll sequence  $t_n = 1$  if  $n > b$ , and  $t_n = 0$  otherwise. Thus  $f_n$  counts the number of nodes that are *not* paged and  $(2^d - 1)f_n + 1$  is the number of pages (external nodes in a tree of arity  $2^d$ ). Theorem 1 applies directly with the Euler transform of the toll sequence satisfying

$$t(z) = \frac{z^{b+1}}{1-z}, \quad t_n^* - t_{n-1}^* = [z^n](-1)^{b+1} \frac{z^{b+1}}{(1-z)^b} = (-1)^{b+1} \binom{n-2}{b-1}.$$

(ii). The asymptotic analysis follows the lines of Theorem 5 as summarized in Figure 1. As evidenced by the figure, the process can be turned into a computer algebra programme to determine the constants  $C^{[b]}$ . We illustrate the computations with the particular case of  $d = 3$  and  $b = 2, 3, 4$ .

The sequence to be extrapolated now involves

$$\nu_k = \sum_{j=b+1}^k \binom{j-2}{b-1} \frac{1}{[j]!},$$

so that an asymptotic analysis of the general term  $u_j = \binom{j-2}{b-1}/[j]!$  is called for. The analytic extrapolation of  $u_j$  is simply the function  $\omega(s)$  given by

$$\omega(s) = \binom{s-2}{b-1} \frac{1}{[s]}.$$

The binomial coefficient is a polynomial of degree  $b-1$  while the ratio  $[+\infty]!/ [j]!$  can be evaluated by means of Stirling's formula applied to the Gamma function form (31). For instance, with  $d = 3$ , one finds

$$\frac{[+\infty]!}{[j]!} \sim 1 - 4\frac{1}{j^2} + 4\frac{1}{j^3} + 6\frac{1}{j^4} - \frac{112}{5}\frac{1}{j^5} + \dots$$

What is needed next is the (easily computable) polynomial  $\alpha(s)$  such that  $\mu_j - \alpha(j)$  tends to 0 as  $j \rightarrow \infty$ . This permits to produce the extrapolating function  $\xi(s)$  in accordance with Lemma 3.

Like for leaves, the function  $[s]!\xi(s)\pi/\sin(\pi s)$  has a simple pole at  $s = 1$ :  $\xi(1) = 0$  as the sum telescopes while  $\pi/\sin(\pi s)$  and  $[s]!$  are singular there. We have:

$$\xi(s) \sim \xi'(1)(s-1); \quad [s]! \sim \frac{2}{d} \frac{1}{s-1}; \quad \frac{\pi}{\sin \pi s} \sim \frac{-1}{s-1}.$$

**Problem.** Given the toll sequence  $t_n$ , find the asymptotic form of the expected cost  $f_n$ .

**Step 1.** Euler transforms and the toll function.

- Determine

$$t(z) = \sum_{n=0}^{\infty} t_n z^n; \quad u(z) = t\left(\frac{z}{z-1}\right); \quad u_n = t_n^* - t_{n-1}^* = [z^n]u(z).$$

**Step 2.** Analytic extrapolation.

- Find  $\omega(s)$  that extrapolates analytically  $u_n$  in the sense that  $u_n = \omega(n)$  for integral values of  $n$ .

- Find  $\xi(s)$  that extrapolates

$$\nu_k = \sum_{j=2}^k \frac{u_j}{[j]!}$$

by the basic formula

$$\xi(s) = \sum_{j=2}^{\infty} \left[ \frac{\omega(j)}{[j]!} - \frac{\omega(j+s-1)}{[j+s-1]!} \right],$$

with the possible addition of convergence terms whenever  $u_j$  does not tend to 0. The convergence terms are produced by an asymptotic analysis of individual terms in accordance with Lemma 3. Thus,  $f_n^*$  is extrapolated by  $[s]! \cdot \xi(s)$ .

**Step 3.** Local analysis of  $f(z)$  near  $z = 1$ .

- Determine  $s_0$ , the pole of largest real part of  $[s]! \xi(s) \pi / \sin(\pi s)$  satisfying  $\Re(s_0) < 2$ , and the constant  $\lambda$  in the singular expansion:

$$[s]! \xi(s) \frac{\pi}{\sin(\pi s)} \sim \frac{\lambda}{(s - s_0)^r}.$$

- Apply Lemmas 1 and 2 to deduce

$$\begin{aligned} f^*(-t) &\sim -\frac{\lambda}{(r-1)!} t^{s_0} (\log t)^{r-1} && \text{as } t \rightarrow +\infty \\ f(z) &\sim -\frac{\lambda}{(r-1)!} (1-z)^{-s_0-1} \left(\log \frac{1}{1-z}\right)^{r-1} && \text{as } z \rightarrow 1. \end{aligned}$$

**Step 4.** Singularity analysis of coefficients.

- Transfer to coefficients by singularity analysis:

$$f_n \sim -\frac{\lambda}{(r-1)! \Gamma(s_0+1)} \frac{n^{s_0}}{(\log n)^{r-1}} \quad \text{as } n \rightarrow \infty.$$

FIGURE 1. The general schema for the asymptotic analysis of quadtree parameters.

This gives the value for the constant  $C^{[b]}$  as

$$\frac{2^{d+1} - 2}{d} \xi'(1).$$

(There is again an extra factor of  $2^d - 1$  because the number of pages is asymptotic to  $(2^d - 1)f_n$  by the conservation law for trees.)

Finally, the value of  $\xi'(1)$  obtains by termwise differentiation of the expression provided by Lemma 3. Here  $n_0 = 2$  and

$$\xi'(s) = A'(s) + \sum_{j=2}^{\infty} \alpha'(s+j-1) - \omega'(s+j-1),$$

so that

$$\xi'(1) = A'(1) + \sum_{j=2}^{\infty} \alpha'(j) - \omega'(j),$$

with  $A(s)$  the summatory function of  $\alpha(s)$ .

For instance, in dimension  $d = 3$ , one finds

$$\begin{aligned} C^{[2]} &= \frac{7}{3} \frac{1}{[+\infty]!} - \frac{14}{3} \sum_{j=2}^{\infty} \left[ \left( \frac{1}{[+\infty]!} - \frac{1}{[j]!} \right) + (j-2) \frac{[\psi](j+1)}{[j]!} \right] \\ C^{[3]} &= -\frac{70}{9} \frac{1}{[+\infty]!} + \frac{7}{3} \sum_{j=2}^{\infty} \left[ (2j-5) \left( \frac{1}{[+\infty]!} - \frac{1}{[j]!} \right) + (j-2)(j-3) \frac{[\psi](j+1)}{[j]!} \right] \\ C^{[4]} &= -\frac{455}{18} \frac{1}{[+\infty]!} - \frac{7}{3} \sum_{j=2}^{\infty} \left[ (2j^2 - 18j + 22) \left( \frac{1}{[+\infty]!} - \frac{1}{[j]!} \right) + (j-2)(j-3)(j-4) \frac{[\psi](j+1)}{[j]!} \right]. \end{aligned}$$

□

When  $d = 1$ , we obtain the exact representation

$$C^{[b]} = \frac{2}{b+2},$$

a result equivalent to the analysis of Quicksort with halting on small subfiles [20, p. 122].

When  $d = 2$ , the explicit form of  $C^{[b]}$  was found in [18],

$$\frac{1}{3} C^{[b]} = 6b^2 + 9b + 1 - 6b(b+1)^2 \left( \frac{\pi^2}{6} - \sum_{j=1}^b \frac{1}{j^2} \right),$$

a result that can also be derived from Theorem 6 thanks to the degeneracy of the  $d$ -Gamma function in this particular case.

When  $d \geq 3$  however, the expression of  $C^{[b]}$  does not lend itself to an easy interpretation, so that it becomes of interest to find simpler numerical approximations. Such approximate forms are suggested by the cases  $d = 1$ , for which  $C^{[b]} \sim 2/b$  and  $d = 3$  where  $C^{[b]} \sim 3/b$ , see [18]. When  $d = 2$  for instance, the formula  $C^{[b]} \sim 3/b$  may be interpreted as representing an equivalent 33% *page filling ratio*: the size of the index tree behaves on average as though pages were 33% of their actual capacity.

**Theorem 7.** *The paging constant  $C^{[b]}$  satisfies*

$$C^{[b]} = \frac{2^{d+1} - d}{d} \cdot \frac{1}{b} \left( 1 + \mathcal{O}\left(\frac{1}{b}\right) \right).$$

*In addition, a full expansion in descending powers of  $b$  holds:*

$$\frac{d}{2^{d+1} - 2} C^{[b]} \sim \frac{1}{b} - \frac{2^d}{3^d - 2^d} \frac{1}{b^2} + \frac{2^d(2^d - 2 \cdot 3^d + 4^d)}{(3^d - 2^d)(4^d - 2^d)} \frac{1}{b^3} + \dots$$

Thus, the equivalent filling ratio is  $d/(2^{d+1} - 2)$  which, for  $d = 1, 2, 3, 4, 5, 6$ , gives

$$\frac{1}{2} \doteq 0.500, \quad \frac{1}{3} \doteq 0.333, \quad \frac{3}{14} \doteq 0.214, \quad \frac{2}{15} \doteq 0.133, \quad \frac{5}{62} \doteq 0.080, \quad \frac{1}{21} \doteq 0.047.$$

The filling ratio decreases steadily as the dimension  $d$ —hence the branching factor—increases, like for other classes of trees, see [25]. As discussed in [18], this can however be compensated by a proper fragmentation of index pages into physical pages of a smaller size, albeit at the expense of increased access time.

**Proof.** The main idea is to analyse *simultaneously* all the paging constants  $C^{[b]}$  by taking a generating function with respect to  $b$  using a generating variable  $w$ . A similar step was taken in [18]. Let  $t_n^{[b]}$  be the toll function of Theorem 6 when the page size is  $b$ . We consider the toll function  $t_n(w) = \sum_b t_n^{[b]} w^b$ , with the generating variable  $w$  restricted, for convergence reasons, to be a parameter that lies in the interval  $(0, 1)$ . The generating function of this toll sequence (with  $w$  a parameter) is

$$t(z) = \sum_n t_n(w) z^n = \frac{z}{(1-z)(1-zw)}.$$

and its Euler transform satisfies

$$t_n^* - t_{n-1}^* = w(1-w)^{n-2}.$$

The analytic extrapolation of the sequence  $t_n^*$  is then simply

$$\omega(s) = w(1-w)^{s-2}.$$

For the asymptotic analysis, we need next to extrapolate

$$[k]! \cdot \nu_k = [k]! \cdot \sum_{j=2}^k \frac{w(1-w)^{j-2}}{[j]!}$$

by

$$[s]! \cdot \xi(s) = [s]! \cdot \sum_{j=2}^{\infty} \left[ \frac{w(1-w)^{j-2}}{[j]!} - \frac{w(1-w)^{j+s-3}}{[j+s-1]!} \right].$$

The dominating singularity of  $[s]! \xi(s) / \sin \pi s$  is at  $s = 1$ . There, we have

$$(38) \quad \xi'(1) = w \log \frac{1}{1-w} \sum_{j=2}^{\infty} \frac{(1-w)^{j-2}}{[j]!} + w \sum_{j=2}^{\infty} \frac{(1-w)^{j-2}}{[j]!} [\psi](j).$$

Like before, the asymptotic form of  $f_n$  (that also involves  $w$  as a parameter) satisfies

$$\lim_{n \rightarrow \infty} \frac{f_n}{n} = \frac{2}{d} \xi'(1).$$

This quantity is none other than the generating function (with  $w$  the generating variable) of the quantities  $C^{[b]}/(2^d - 1)$ :

$$(39) \quad C(w) := \sum_{b=1}^{\infty} C^{[b]} w^b = \frac{2^{d+1} - 2}{d} \xi'(1).$$

The expression of  $\xi'(1)$  in (38) gives a representation of  $C(w)$  converging not only for  $w \in (0, 1)$  but even for all  $|w - 1| < 1$ . In other words, the expansion of  $\xi'(1)$  furnishes the singular expansion of  $C(w)$  near  $w = 1$ . The first sum in (38) has a logarithmic factor. The second sum is analytic at  $w = 1$ . Thus,  $C(w)$  has a logarithmic singularity at  $w = 1$  whose behaviour is described by the first of the two sums.

Now, by the principles of singularity analysis, a termwise translation from the singular expansion of  $C(w)$  to its coefficients is permitted. Taking care of the extra factor of  $2^d - 1$  and translating (38) by the rule

$$[z^b] w(1-w)^r \log \frac{1}{1-w} = \frac{(-1)^r r!}{(b-1)(b-2)\cdots(b-r)},$$

one finds

$$\frac{d}{2^{d+1} - 2} C^{[b]} \sim \sum_{r=0}^{\infty} \frac{(-1)^r r!}{(b-1)(b-2)\cdots(b-r-1)} \frac{1}{[r+2]}.$$

This expansion can in turn be converted into a standard expansion into descending powers of  $b$ .

In particular, for  $d = 1, 2, 3, 4$ , one gets

$$(40) \quad \begin{aligned} d = 1 & : \frac{1}{2} C^{[b]} \sim \frac{1}{b} - \frac{2}{b^2} + \frac{4}{b^3} - \frac{8}{b^4} + \frac{16}{b^5} + \cdots \\ d = 2 & : \frac{1}{3} C^{[b]} \sim \frac{1}{b} - \frac{4}{5b^2} + \frac{2}{5b^3} + \frac{2}{35b^4} - \frac{2}{7b^5} + \cdots \\ d = 3 & : \frac{3}{14} C^{[b]} \sim \frac{1}{b} - \frac{8}{19b^2} - \frac{2}{133b^3} + \frac{226}{1729b^4} + \frac{562}{22477b^5} + \cdots \\ d = 4 & : \frac{2}{15} C^{[b]} \sim \frac{1}{b} - \frac{16}{65b^2} - \frac{2}{25b^3} + \frac{2822}{65975b^4} + \frac{2}{29b^5} + \cdots \end{aligned}$$

For  $d = 1$ , we recover the expansion of  $1/(b + 2)$ . For  $d = 2$ , this series coincides —as expected— with the one obtained in [18]. These expressions give excellent numerical approximations, even for small values of  $b$  and they thus effectively complement the exact representations of Theorem 6. Already with  $d = 2$ , the relative error provided by the expansions of (40) truncated after their  $b^{-5}$  terms is

$$7 \cdot 10^{-3}, \quad 2 \cdot 10^{-4}, \quad 3 \cdot 10^{-6}, \quad 8 \cdot 10^{-8},$$

for  $b = 2, 4, 8, 16$  respectively. □

## 6. NODE LEVELS AND EXTREME POINTS

In this section, we examine two computationally related problems: the distribution of node depths in randomly grown quadtrees and the analysis of the algorithm that finds extreme points in quadtrees.

**Node levels.** Let  $\tau_{n,k}$  denote the mean number of internal nodes at depth  $k + 1$  in a randomly grown quadtree of size  $n$  (by convention the root is taken to be at depth 1). The probability that the search of a random element stored in the tree (*i.e.*, a positive search) requires  $k$  node comparisons is  $\tau_{n,k-1}/n$ . Thus, by computing these averages parameterized by  $k$ , we are analysing the probability distribution of the cost of a random positive search.

The generating polynomial

$$\tau_n(u) = \sum_{k=0}^n \tau_{n,k} u^k$$

is a solution to the recurrence

$$f_n = 1 + 2^d u \sum_{k=0}^{n-1} \pi_{n,k} f_k \quad \text{with} \quad f_0 = 0, f_1 = 1.$$

In dimension 1, the  $\tau_n(u)$  are closely related to the Stirling polynomials of the first kind, a result of Lynch, see [20, Ex. 6.2.6]. In dimension 2, the  $\tau_{n,k}$  have been expressed in [9] as convolutions of Stirling numbers but no exact expression was previously known for  $d \geq 3$ . We have:

**Theorem 8.** *The level polynomials are given by*

$$\tau_n(u) = nu - u(1-u) \sum_{k=2}^n \binom{n}{k} (-1)^k [k; u]!$$

where

$$[k; u]! = \left(1 - u \frac{2^d}{3^d}\right) \cdots \left(1 - u \frac{2^d}{k^d}\right), \quad [2; u]! = 1.$$

**Proof.** The proof mimics the derivation of the algebraic solution of the quadtree recurrence based on the Euler transformation of Section 2. We find:

$$f_n = \sum_k \binom{n}{k} (-1)^k f_k^*, \quad f_n^* = -\delta_{n,1} + \left(1 - u \frac{2^d}{n^d}\right) f_{n-1}^*.$$

□

The Stirling polynomials of the first kind are classically defined by [1]

$$s_n(u) = u(u+1)(u+2) \cdots (u+n-1).$$

By factoring the general term of  $[k; u^d]!$ , one finds

$$[k; u^d]! = \left(\frac{2}{k!}\right)^d \prod_{\omega^d=2^d} \widehat{s}_n(-u\omega) \quad \text{where} \quad \widehat{s}_n(u) = \frac{s_{n+1}(u)}{u(u+1)(u+2)}.$$

Thus the level polynomials are expressible in terms of Stirling polynomials in all dimensions. Their coefficients are in turn expressible as complicated convolutions of Stirling numbers. As the Stirling numbers admit an elementary expression by Schömlich's formula [1], the probability distribution of search costs is in principle even reducible to a (multiple!) summation of entirely elementary functions.

Flajolet and Lafforgue [9] have proved that the variance is asymptotic to  $\frac{2}{d} \log n$  (it would be possible to derive a complicated expression by taking double derivatives), from which follows that the depth of a random node  $D_n$  satisfies  $\frac{D_n}{\log n} \rightarrow \frac{2}{d}$  in probability, an earlier result of Devroye and Laforest proved by probabilistic arguments in [4]. The fact that the distribution of  $D_n$  is Gaussian

in the asymptotic limit [9] could probably be derived by the asymptotic techniques of the present paper via an analysis of the  $s_n(u)$  for  $u$  in a complex neighbourhood of 1 (see Theorem 8 for the particular case of  $u = 1/2$ ), combined with continuity theorems for characteristic functions.

**Extreme points.** The recurrence for the mean number of node comparisons in the search for points with smallest  $x$ -coordinate is [8]

$$f_n = 1 + 2^{d-1} \sum_{k=0}^{n-1} \pi_{n,k} f_k, \quad f_0 = 0.$$

This happens to be a particular case of the recurrence for level polynomials, so that  $f_n = s_n(\frac{1}{2})$ , with previous notations. The asymptotics derives from the schema described in Figure 1.

**Theorem 9.** *The expected number of node comparisons in the search of extreme points is*

$$f_n = \frac{1}{2}n - \frac{1}{4} \sum_{k=2}^n \binom{n}{k} (-1)^k [k; \frac{1}{2}]!$$

where

$$[k; \frac{1}{2}]! = \left(1 - \frac{2^{d-1}}{3^d}\right) \cdots \left(1 - \frac{2^{d-1}}{k^d}\right), \quad [2; \frac{1}{2}]! = 1.$$

Asymptotically, as  $n \rightarrow +\infty$ ,

$$f_n \sim E n^{2^{(d-1)/d} - 1}$$

where the constant  $E$  is rationally expressible in terms of Gamma functions of algebraic arguments:

$$E = -2^{d-2} \omega_0 \frac{\Gamma(-\omega_0)}{\Gamma(\omega_0)^d} \frac{\prod'_\omega \Gamma(\omega_0 - \omega)}{\prod_\omega \Gamma(3 - \omega)}.$$

There  $\omega_0 = 2^{(d-1)/d}$ ,  $\prod_\omega$  represents a product over all roots of  $\omega^d = 2^{d-1}$  and  $\prod'_\omega$  the product over all such roots  $\omega$  with  $\omega \neq \omega_0$ .

For  $d = 2, 3, 4$ , the growth rate is thus

$$n^{\sqrt{2}-1} \approx n^{0.41}, \quad n^{2^{2/3}-1} \approx n^{0.58}, \quad n^{2^{3/4}-1} \approx n^{0.68}.$$

The general asymptotic pattern was found in [8], though no explicit form was known for the constant  $E$  as soon as  $d > 2$ . The reader is invited to correct two typographical errors in the statement of Theorem 6 of [8]: for  $d \geq 2$ , the exponent should read  $\alpha = 2^{(d-1)/d} - 1$  (in accordance with Theorem 9 above) and for  $d = 2$ , the value of the constant  $E$  is

$$E = \frac{\Gamma(2\sqrt{2})}{2\sqrt{2}(\Gamma(\sqrt{2}))^3} \doteq 0.87126.$$



## 7. CONCLUSIONS

First, a few words about methodology. It came somewhat as a surprise that the Euler transform would drastically simplify the analysis of quadtrees by reducing recurrences to order 1. It is a coincidence that the analysis of digital search trees and their generalizations also simplifies under this transformation [11].

The method of integral representations *a la* Mellin-Lindelöf is probably the one of widest potential scope. We have applied it here in the context of alternating binomial sums. As noted by Messrs. Gourdon and Prodinger, such sums also fall into the range of “Rice’s integrals” commonly used for the analysis of digital search trees, see [20] and [12] for a survey. The “Mellin-Lindelöf” method is however in no way restricted to the asymptotic analysis of Euler transforms.

Next, we briefly summarize what is known at present about quadtree parameters.

The mean search cost is  $\sim (2/d)\log n + \mu_d$ , with  $\mu_d$  given by Theorem 3; the variance is  $\sim (2/d^2)\log n$  as shown in [4] for  $d = 2$  and in [9] for all  $d$ ; this implies the convergence-in-probability result of [4]; in addition, the distribution of search costs is asymptotically normal with exponential tails [9].

The height presents a different problem solved in [2, 3] and reviewed in [25]. The expected height is still logarithmic and asymptotic to  $\frac{2c}{d}\log n$  where  $c = 4.31107$ . As far as we know the problems of estimating the variance and the limit distribution of height are still open.

Partial match queries with  $s$  out of  $d$  coordinates specified have expected cost  $\sim cn^{\alpha-1}$  where  $\alpha$  is a well-characterized algebraic number of degree  $d$ , see [8]. The constant  $c$  is only known in closed form when  $d = 2$  (and  $s = 1$ ), thanks to explicit  ${}_2F_1$ -hypergeometric forms available in this particular case. It is not clear whether methods of the present paper might give access to the constants involved when  $d \geq 3$ . However, the search for extreme points is completely solved by Theorem 9.

Theorems 4 and 5 provide a precise description of the number of leaves in quadtrees. Statistics of other node types for  $d = 2$  are discussed by Labelle-Laforest in [22, 21, 23] where it is proved that the proportion of nodes with 0,1,2,3,4 nonempty children is respectively

$$\begin{aligned} \lambda_0 &= 4\pi^2 - 39, & \lambda_1 &= 24\zeta(3) - 26\pi^2 + 228, \\ \lambda_2 &= -132\zeta(3) + 24\pi^2\log(2) + \frac{67}{2}\pi^2 - 336, & \lambda_3 &= 192\zeta(3) - 48\pi^2\log(2) - 5\pi^2 + 147, \\ \lambda_4 &= 1 - 84\zeta(3) + 24\pi^2\log(2) - \frac{13}{2}\pi^2. \end{aligned}$$

It would be of interest to extend the analysis to higher dimensions.

A fairly complete discussion of paging for dimension  $d = 2$  appears in [18] where the approximate formula  $3n/b$  was first obtained. Theorems 6, 7 demonstrate that higher dimensions can be treated by our methods. In this context, the most significant result derives from a combination of Theorem 6 and Theorem 7 which provide an approximate formula for the number of pages in a quadtree formed with  $n$  points when the page capacity equals  $b$ :

$$(41) \quad \Pi(n, b) = \left( \frac{2^{d+1} - 2}{d} \right) \cdot \binom{n}{b}.$$

(This is obtained by first taking the limit for large  $n$  in Theorem 6 and then using the approximate form of the paging constants of Theorem 7.) The approximation is excellent, as evidenced by

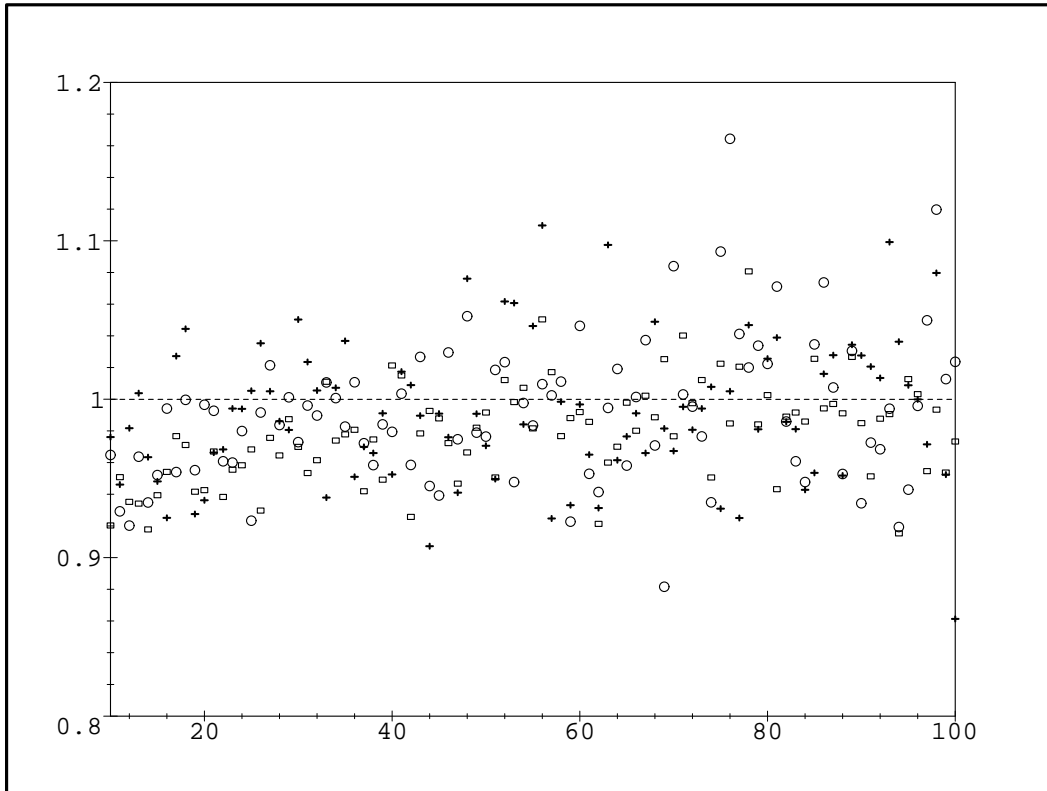


FIGURE 2. Paged quadtrees. A display of the ratio between the number of pages observed by simulation and the approximate formula  $\Pi(n, b)$  of (41). The value of  $n$  is fixed to  $n = 10,000$ . The page capacity varies from  $b = 10$  to  $b = 100$ . Each dot represents one simulation in dimension  $d = 2$  (rectangles),  $d = 3$  (circles) and  $d = 4$  (crosses).

Figure 2 which displays the ratio between values observed from simulations and the approximation  $\Pi(n, b)$ : in each of the 273 simulations, the ratio between observed values and predicted values always lies between 0.85 and 1.16. As explained in [18], this may be used to configure algorithms with well quantified page filling versus page access trade-offs.

#### APPENDIX

In this section, we recall various forms of the splitting probabilities for the quadtree process taken from [8, 22, 23].

**Lemma 4.** *The probability that the first root subtree of a quadtree of size  $n$  has size  $k$  admits the following expressions:*

(i) *Multiple sums:*

$$\pi_{n,k} = \frac{1}{n} \sum_{k < i_1 \leq i_2 \leq \dots \leq i_{d-1} \leq n} \frac{1}{i_1 \cdots i_{d-1}}.$$

(ii) *Expressions in terms of generalized harmonic numbers  $H_m^{(r)} = \sum_{j=1}^m j^{-r}$ :*

$$\pi_{n,k} = \frac{1}{n} \sum_{n_1+2n_2+\dots+(d-1)n_{d-1}=d} \frac{(H_n - H_k)^{n_1} \cdots (H_n^{(d-1)} - H_k^{(d-1)})^{n_{d-1}}}{1^{n_1} n_1! \cdots (d-1)^{n_{d-1}} n_{d-1}!}.$$

(iii) *Alternating binomial sums:*

$$\pi_{n,k} = \binom{n-1}{k} \sum_{i=1}^{n-1-k} \binom{n-1-k}{i} \frac{(-1)^i}{(k+i+1)^d}.$$

**Proof.** The basic form (i) is proved in [8] by a direct reduction to Eulerian beta integrals. From this form, it is also apparent that

$$\pi_{n,k} = \frac{1}{n} [z^{d-1}] \prod_{i=k+1}^n \frac{1}{1-z/i},$$

and the product can be expanded as

$$\prod_{i=k+1}^n \frac{1}{1-z/i} = \exp \left( \sum_{i=k+1}^n \log(1-z/i)^{-1} \right) = \exp \left( \sum_{n=1}^{\infty} (H_n^{(r)} - H_k^{(r)}) z^n \right),$$

whence the occurrence of generalized harmonic numbers in (ii).

The last form (iii) results from

$$\pi_{n,k} = \binom{n-1}{k} \int_0^1 \cdots \int_0^1 (u_1 \cdots u_d)^k (1-u_1 \cdots u_d)^{n-1-k} du_1 \cdots du_d,$$

which estimates the probability by first conditioning upon the fact that the root has its  $i$ th coordinate in the interval  $[u_i, u_i + du_i]$ . Form (iii) derives by expanding the integrand and integrating termwise. A form equivalent to (iii) is

$$\pi_{n,k} = \binom{n-1}{k} \int_0^1 t^k (1-t)^{n-1-k} \frac{(-\log t)^{d-1}}{(d-1)!} dt,$$

which may be checked again by direct integration. □

## REFERENCES

1. COMTET, L. *Advanced Combinatorics*. Reidel, Dordrecht, 1974.
2. DEVROYE, L. A note on the expected height of binary search trees. *Journal of the ACM* 33 (1986), 489–498.
3. DEVROYE, L. Branching processes in the analysis of the heights of trees. *Acta Informatica* 24 (1987), 277–298.
4. DEVROYE, L., AND LAFOREST, L. An analysis of random  $d$ -dimensional quad trees. *SIAM Journal on Computing* 19 (1990), 821–832.
5. DINGLE, R. B. *Asymptotic Expansions: Their Derivation and Interpretation*. Academic Press, London, 1973.
6. ERDÉLYI, A. *Higher Transcendental Functions*, second ed., vol. 1-2-3. R. E. Krieger publishing Company, Inc., Malabar, Florida, 1981.
7. FINKEL, R. A., AND BENTLEY, J. L. Quad trees, a data structure for retrieval on composite keys. *Acta Informatica* 4 (1974), 1–9.
8. FLAJOLET, P., GONNET, G., PUECH, C., AND ROBSON, J. M. Analytic variations on quadtrees. *Algorithmica* 10, 7 (Dec. 1993), 473–500.
9. FLAJOLET, P., AND LAFFORGUE, T. Search costs in quadtrees and singularity perturbation asymptotics. *Discrete and Computational Geometry* 12, 4 (1994), 151–175.
10. FLAJOLET, P., AND ODLYZKO, A. M. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics* 3, 2 (1990), 216–240.
11. FLAJOLET, P., AND RICHMOND, B. Generalized digital trees and their difference–differential equations. *Random Structures and Algorithms* 3, 3 (1992), 305–320.
12. FLAJOLET, P., AND SEDGEWICK, R. Mellin transforms and asymptotics: finite differences and Rice’s integrals. Research Report 2031, Institut National de Recherche en Informatique et en Automatique, Mar. 1994. 21 pages. To appear in *Theoretical Computer Science*.
13. FORD, W. B. *Studies on divergent series and summability and the asymptotic developments of functions defined by Maclaurin series*, 3rd ed. Chelsea Publishing Company, New York, 1960.
14. GASPER, G., AND RAHMAN, M. *Basic Hypergeometric Series*, vol. 35 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1990.
15. GONNET, G. H. *Handbook of Algorithms and Data Structures*. Addison–Wesley, 1984.
16. GONNET, G. H., AND BAEZA-YATES, R. *Handbook of Algorithms and Data Structures: in Pascal and C*, second ed. Addison–Wesley, 1991.
17. HARDY, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by his Life and Work*, third ed. Chelsea Publishing Company, New-York, 1978. Reprinted and Corrected from the First Edition, Cambridge, 1940.
18. HOSHI, M., AND FLAJOLET, P. Page usage in a quadtree index. *BIT* 32 (1992), 384–402.
19. IWASAKI, K., KIMURA, H., SHIMOMURA, S., AND YOSHIDA, M. *From Gauss to Painlevé—A Modern Theory of Special Functions*. Aspects of Mathematics. Vieweg, Braunschweig, 1991.
20. KNUTH, D. E. *The Art of Computer Programming*, vol. 3: Sorting and Searching. Addison-Wesley, 1973.
21. LABELLE, G., AND LAFOREST, L. Étude de constantes universelles pour les arborescences hyperquaternaires de recherche, 1993. Proceedings of the 5th FPSAC conference, Florence, June 1993.
22. LABELLE, G., AND LAFOREST, L. Combinatorial variations on multidimensional quadtrees, 1994. *Journal of Combinatorial Theory*, Series A, to appear.
23. LAFOREST, L. Étude des arbres hyperquaternaires. Tech. Rep. 3, LACIM, UQAM, Montreal, Nov. 1990. (Author’s PhD Thesis at McGill University).
24. LINDELÖF, E. Le calcul des résidus et ses applications à la théorie des fonctions. In *Collection de monographies sur la théorie des fonctions, publiée sous la direction de M. Émile Borel*. Gauthier-Villars, Paris, 1905. Reprinted by Gabay, Paris, 1989.
25. MAHMOUD, H. *Evolution of Random Search Trees*. John Wiley, New York, 1992.
26. SAMET, H. *Applications of Spatial Data Structures*. Addison–Wesley, 1990.

27. SAMET, H. *The Design and Analysis of Spatial Data Structures*. Addison–Wesley, 1990.
28. SLATER, L. J. *Generalized Hypergeometric Functions*. Cambridge University Press, 1966.
29. WHITTAKER, E. T., AND WATSON, G. N. *A Course of Modern Analysis*, fourth ed. Cambridge University Press, 1927. Reprinted 1973.

Ph. Flajolet: Algorithms Project, INRIA, Rocquencourt, F–78153 Le Chesnay, France.

*E-mail address*: Philippe.Flajolet@inria.fr

G. Labelle: Département de Mathématiques et d’Informatique, Université du Québec à Montréal, Case Postale 8888, Succursale centre-ville, Montréal, Quebec H3C3P8 Canada.

*E-mail address*: gilbert@lacim.uqam.ca

L. Laforest: Département de Mathématiques et d’Informatique, Université du Québec à Montréal, Case Postale 8888, Succursale centre-ville, Montréal, Quebec H3C3P8 Canada.

*E-mail address*: laforest@lacim.uqam.ca

B. Salvy: Algorithms Project, INRIA, Rocquencourt, F–78153 Le Chesnay, France.

*E-mail address*: Bruno.Salvy@inria.fr