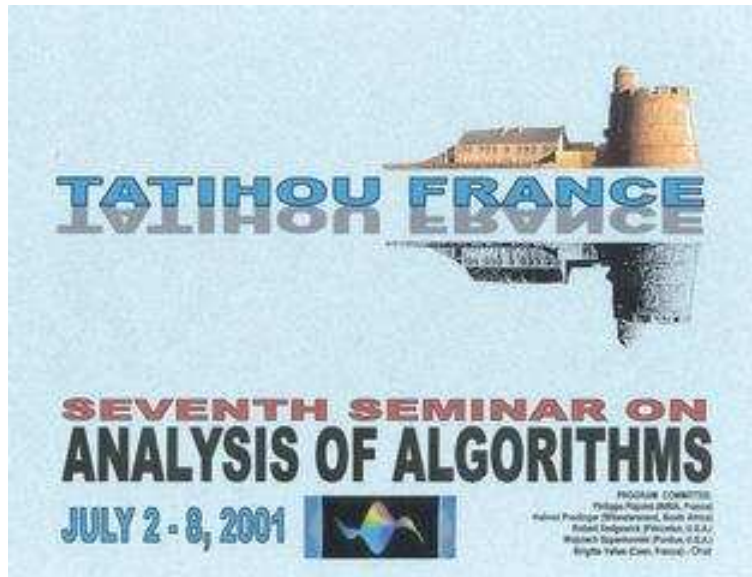


The Seventh Seminar on *Analysis of Algorithms*

~~~~~ AofA'2001 ~~~~~

1-7 July 2001, Tatihou (France)



## Abstracts

|                                                                                                  |    |
|--------------------------------------------------------------------------------------------------|----|
| <b>Cyril Banderier</b> .....                                                                     | 4  |
| <i>Combinatorial and analytic properties of lattice paths</i>                                    |    |
| <b>Brigitte Chauvin</b> .....                                                                    | 5  |
| <i>Discrete martingales and a few applications including binary search trees</i>                 |    |
| <b>Hua-Huai, Felix, Chern</b> .....                                                              | 6  |
| <i>An elementary approach to the asymptotics of quicksort-type recurrences with applications</i> |    |
| <b>Kevin J. Compton</b> .....                                                                    | 7  |
| <i>An Analysis Inspired by Gosper's Algorithm</i>                                                |    |
| <b>Luc Devroye</b> .....                                                                         | 8  |
| <i>Tries</i>                                                                                     |    |
| <b>Michael Drmota</b> .....                                                                      | 9  |
| <i>The height of digital search trees</i>                                                        |    |
| <b>James A. Fill</b> .....                                                                       | 10 |
| <i>Singularity analysis and the probabilistic analysis of a class of search-tree functionals</i> |    |
| <b>Philippe Flajolet</b> .....                                                                   | 11 |
| <i>Algebraic Analytic Asymptotics</i>                                                            |    |
| <b>Hsien-Kuei Hwang</b> .....                                                                    | 12 |
| <i>A method of moments for random recursive structures</i>                                       |    |
| <b>Svante Janson</b> .....                                                                       | 13 |
| <i>Quicksort asymptotics</i>                                                                     |    |
| <b>Guy Louchard</b> .....                                                                        | 14 |
| <i>Reflected Brownian Bridge area conditioned on its local time at the origin</i>                |    |
| <b>Hosam M. Mahmoud</b> .....                                                                    | 15 |
| <i>The Size of Random Bucket Trees via Urn Models</i>                                            |    |
| <b>Jean-François Marckert</b> .....                                                              | 16 |
| <i>Non-Crossing trees are almost conditioned Galton-Watson trees</i>                             |    |
| <b>Conrado Martínez</b> .....                                                                    | 17 |
| <i>On the complexity of unranking and ordered generation</i>                                     |    |
| <b>Donatella Merlini</b> .....                                                                   | 18 |
| <i>Some matrices often arising in combinatorics and in the analysis of algorithms</i>            |    |
| <b>Markus E. Nebel</b> .....                                                                     | 19 |
| <i>RNA Secondary Structures and Their Relation to Binary Trees</i>                               |    |
| <b>Ralph Neininger</b> .....                                                                     | 20 |
| <i>Multivariate Contraction Method</i>                                                           |    |
| <b>Michel Nguyen-The</b> .....                                                                   | 21 |
| <i>Distributions of Valuations on Trees</i>                                                      |    |
| <b>Daniel Panario</b> .....                                                                      | 22 |
| <i>Pairs of coprime smooth polynomials and the Waterloo algorithm</i>                            |    |

|                                                                                           |    |
|-------------------------------------------------------------------------------------------|----|
| <b>Jordi Petit</b> .....                                                                  | 23 |
| <i>Hamiltonian Cycles in Faulty Random Geometric Networks</i>                             |    |
| <b>Boris Pittel</b> .....                                                                 | 24 |
| <i>Phase transition and finite-size scaling for the integer partitioning problem</i>      |    |
| <b>Nicolas Pouyanne</b> .....                                                             | 25 |
| <i>Permutations admitting an <math>m</math>-th root</i>                                   |    |
| <b>Helmut Prodinger</b> .....                                                             | 26 |
| <i>Randomized search of fixed length binary words - or - a trie model of two russians</i> |    |
| <b>Mireille Regnier</b> .....                                                             | 27 |
| <i>Large Deviations on Words</i>                                                          |    |
| <b>Ludger Rueschendorf</b> .....                                                          | 28 |
| <i>A general limit theorem for recursive algorithms</i>                                   |    |
| <b>Gilles Schaeffer</b> .....                                                             | 29 |
| <i>Random planar maps and alternating knots and links</i>                                 |    |
| <b>Hadas Shachnai</b> .....                                                               | 30 |
| <i>Efficient Reorganization of Binary Search Trees</i>                                    |    |
| <b>Renzo Sprugnoli</b> .....                                                              | 31 |
| <i>Histograms and Fountains</i>                                                           |    |
| <b>Wojciech Szpankowski</b> .....                                                         | 32 |
| <i>New and Old Problems in Pattern Occurrences</i>                                        |    |
| <b>Brigitte Vallée</b> .....                                                              | 33 |
| <i>Hidden pattern statistics</i>                                                          |    |
| <b>Alfredo Viola</b> .....                                                                | 34 |
| <i>Analysis of Rabin's Irreducibility Test for Polynomials over Finite Fields</i>         |    |
| <b>Andreas Weiermann</b> .....                                                            | 35 |
| <i>Analytic combinatorics and rapidly growing functions</i>                               |    |

**Cyril Banderier**

Projet Algo, INRIA Rocquencourt  
Cyril.Banderier@inria.fr

## **Combinatorial and analytic properties of lattice paths**

GF of walks (lattice paths) over  $\mathbb{N}$  and GF of walks over  $\mathbb{Z}$  are related by nice combinatorial formulae (related to identities known in probability theory under the name of Spitzer formula), reflecting a variation of the cyclic lemma.

They rely on the theory of the  $\Theta$  (pointing) operator and on some peculiar factorization (e.g. an excursion can be factorized into arches).

This gives access to some parameters, such as the number of times that the minimum is reached, for which I derive the asymptotics.

In general, there exist 3 distinct cases (depending on the "drift" of the walk) leading to 3 distinct limit laws.

A deep use is made of the kernel method, our asymptotic analysis (saddle point method and analysis of singularities) relies indeed on the analytic properties of the roots of the kernel.

(Work related to my PhD thesis and a paper in preparation with Ph. Flajolet<sup>1</sup>).

---

---

<sup>1</sup>INRIA Rocquencourt, Philippe.Flajolet@inria.fr

**Brigitte Chauvin**

Département de Mathématiques, Université de Versailles Saint-Quentin  
chauvin@math.uvsq.fr

## **Discrete martingales and a few applications including binary search trees**

In a first part, basic notions of the discrete martingale theory are given: decompositions, inequalities, stopping theorems and main convergence theorems. Several examples allow to see how martingales can be found, how a martingale can be transformed and how a martingale concentrates the randomness. The case of exponential martingales is studied: first, it gives an easy example of a non  $L^1$  convergent, nonnegative martingale. They are also a basic tool in the study of branching random walks (which are related to the height of a binary search tree). Moreover, exponential martingales provide the tails of the distributions of sums of random variables. On the same topics, namely the concentration of a random variable around its mean, the Hoeffding-Azuma inequality is mentioned and it is applied in the bin-packing problem.

In a second part, the previous tools are used for binary search trees. The convergence theorems give information on some cost variables. The random measure valued

$$\nu_n = \sum_{k \geq 0} U_k(n) \delta_k,$$

where  $U_k(n)$  is the number of nodes at level  $k$  in the tree with  $n$  nodes), describes the repartition of the nodes in the tree and it is studied here. The multidimensional martingales are mentioned and used to get convergence of  $C$ -valued random variables, for instance the "level polynomial"

$$W_n(z) = \sum_{k \geq 0} U_k(n) z^k$$

Finally, extensions to the  $m$ -ary search trees are given.

---

Hua-Huai, Felix, Chern

Department of Mathematics and Science Educations, Taipei Municipal Teachers College  
felix@mail1.tmtc.edu.tw

## An elementary approach to the asymptotics of quicksort-type recurrences with applications

*This is joint work with Hsien-Kuei Hwang<sup>1</sup> and Tsung-Hsi Tsai<sup>2</sup>.*

The average-case analysis of sorting algorithms and many random data structures often necessitates the resolution of some recurrences, linear or nonlinear. In this talk I will introduce a simple elementary approach (without complex analysis) to the “asymptotic transfers” for quantities defined by a recurrence equation of quite general type. The approach reveals not only the intrinsic insight of such recurrences but also reflects more analytic properties of the associated PGF and differential equations. Almost no knowledge on differential equations is needed for this approach. The class of recurrences we study cover examples like the generalized quicksort of Hennequin that has recently been analyzed by purely analytic approach, the analysis of quicksort using Tukey’s “ninther” (the median of three medians-of-three) and its extensions. Our approach can be used to derive all moments (centralized or not) and limit laws of the main cost measures in a systematic way.

The interest of studying “ninther” is multi-fold: first the underlying cost measures introduce naturally several new recurrences of quicksort-type that are so messy that a more algebraic and abstract treatment is needed; second, we can vary the underlying scheme to produce more instances of “phase changes” of the limit laws.

---

---

<sup>1</sup>Institute of Statistical Science, Academia Sinica, [hkhwang@stat.sinica.edu.tw](mailto:hkhwang@stat.sinica.edu.tw)

<sup>2</sup>Institute of Statistical Science, Academia Sinica, [chonghi@stat.sinica.edu.tw](mailto:chonghi@stat.sinica.edu.tw)

Kevin J. Compton

EECS Department, University of Michigan  
kjc@umich.edu

## An Analysis Inspired by Gosper's Algorithm

*This is joint work with Olga Milenkovic<sup>1</sup>.*

Gosper's algorithm (*GA*) produces, for a given hypergeometric term  $t(n)$ , another hypergeometric term  $z(n)$  such that  $t(n) = z(n+1) - z(n)$ . If there is no such  $z(n)$ , *GA* determines that there is no solution. If *GA* succeeds, the sum of the terms  $t(n)$  over  $n = 0, \dots, k-1$  is simply  $z(k) - z(0)$ . By definition,  $t(n+1)/t(n)$  is a rational function; write it as a ratio  $f(n)/g(n)$ , where  $f$  and  $g$  are relatively prime polynomials. A basic step in *GA* is to rewrite this ratio in the form

$$\frac{c(n+1)a(n)}{c(n)b(n)},$$

where  $a$ ,  $b$  and  $c$  are polynomials with  $a(n)$  and  $b(n+h)$  relatively prime for all nonnegative integers  $h$ . The degree of  $c(n)$  can be quite large and (in most cases) controls the running time of *GA*. We determine the expected degree of  $c(n)$  for two probabilistic models. In the first model  $f(n)$  and  $g(n)$  are products of the random factors of the form  $(n-a)$  where  $a$  is chosen uniformly from  $\{1, 2, \dots, m\}$  (with repetitions). This is the "ball and urn" model. In the second model, the set of roots of  $f(n)$  (and  $g(n)$ ) are random multisets chosen uniformly from a set of multisets of a given size. For the first model, we use Poissonization. For the second, we use a similar method based on the geometric distribution rather than the Poisson distribution.

---

---

<sup>1</sup>EECS Department, University of Michigan

**Luc Devroye**

School of Computer Science, Mc Gill University, Montreal  
luc@cs.mcgill.ca

## **Tries**

We revisit tries from several perspectives. Topics include universal laws of large numbers for trie parameters, partial match, and level compaction.

---

**Michael Drmota**

Institut für Geometrie, Technische Universität Wien  
michael.drmota@tuwien.ac.at

## The height of digital search trees

It is shown that the expected value of the height  $H_n$  of digital search trees is given by

$$\mathbf{E} H_n = \log_2 n + \sqrt{2 \log_2 n} (1 + o(1))$$

and that the variance is bounded:

$$\mathbf{Var} H_n = O(1).$$

This result follows by analyzing generating functions  $Y_n(z)$  satisfying the recurrent relation  $Y'_{n+1}(z) = Y_n(z/2)^2$ .

---

**James A. Fill**

Department of Mathematical Sciences, The Johns Hopkins University, Baltimore  
jimfill@jhu.edu

## **Singularity analysis and the probabilistic analysis of a class of search-tree functionals**

*This is a joint work with Philippe Flajolet<sup>1</sup> and Nevin Kapur<sup>2</sup>.*

We review the notion of an additive-type functional on  $m$ -ary search trees. Three examples of such functionals are space requirement, total path length, and log-product of branch sizes; the last of these three serves as a crude measure of the "shape" of the tree. The goal is to study the distribution of an additive-type functional applied to a tree built in a natural way from a uniformly random permutation of  $\{1, \dots, n\}$ .

We provide a general framework for the exact and asymptotic analysis of such distributions. Singularity analysis (the extraction of asymptotic information about a sequence from the behavior of its generating function near singularities) can be extremely useful in the consideration of asymptotic distributions, but the tools currently available do not handle the log-product functional. We expand the singularity analysis tool kit by proving that if singularity analysis can be applied to each of two sequences, then it can also be applied to the Hadamard product of the sequences. This tool allows for the handling not only of the log-product functional, but also of a wide variety of asymptotic problems in combinatorics and probability.



---

<sup>1</sup>INRIA-Rocquencourt, France, [Philippe.Flajolet@inria.fr](mailto:Philippe.Flajolet@inria.fr)

<sup>2</sup>The Johns Hopkins University, [kapur@mts.jhu.edu](mailto:kapur@mts.jhu.edu)

**Philippe Flajolet**

Projet Algo, INRIA Rocquencourt

Philippe.Flajolet@inria.fr

## Algebraic Analytic Asymptotics

Many combinatorial problems lead to algebraic generating functions, that is, functions that are solutions of polynomial equations. Asymptotic enumeration results then depend on being able to determine precisely the location and nature of singularities. The major theorem of “Drmotá-Lalley-Woods” addresses such questions in the case of an important but special class of algebraic functions (the positive ones arising from “strongly dependent” polynomial systems).

The difficulty of the problem, at its fullest level of generality stems from the fact that algebraic functions are inherently multivalued. There appears consequently a “connection problem”: Given initial conditions dictated by combinatorics, where are and which are the “relevant” singularities? We prove this problem to be decidable in relatively low computational complexity. A consequence is the effective computability of coefficient asymptotics for general algebraic functions.

The results are potentially applicable to any combinatorial class that admits of an algebraic generating function. Examples (to be evoked, time permitting) include all combinatorial classes described by context-free languages (via the Chomsky-Schützenberger Theorem), several random walks and polyomino problems (via the “kernel method”), several geometric configuration problems (e.g., the noncrossing configurations studied by Flajolet-Noy), trees and terms (following Meir and Moon’s framework), planar maps (via Tutte’s quadratic method), and so on.

[Some of this research is described in “Analytic combinatorics: functional equations, rational, and algebraic functions” by P. Flajolet, R. Sedgewick<sup>1</sup> (Res. Rep. INRIA RR4103, January 2001, 98 pages). Some results represent ongoing research with Bruno Salvy<sup>2</sup> and Cyril Chabaud<sup>3</sup>.]

---

---

<sup>1</sup>Department of Computer Science, Princeton University, [rs@cs.princeton.edu](mailto:rs@cs.princeton.edu)

<sup>2</sup>Projet Algo, INRIA Rocquencourt, [Bruno.Salvy@inria.fr](mailto:Bruno.Salvy@inria.fr)

<sup>3</sup>Projet Algo, INRIA Rocquencourt, [Cyril.Chabaud@inria.fr](mailto:Cyril.Chabaud@inria.fr)

**Hsien-Kuei Hwang**

Institute of Statistical Science, Academia Sinica  
hkhwang@stat.sinica.edu.tw

## **A method of moments for random recursive structures**

A systematic approach for limit laws based on calculating the asymptotics of higher moments is presented. When applied to recursively defined random variables, the approach reduces the calculations of all moments to essentially the derivation of "asymptotic transfers," which bridge the asymptotics of the costs of subproblems to that of the total cost in the underlying recurrence of the moments. Many applications of this approach to the analysis of algorithms will be indicated, including maxima in right triangle, quickselect, maximum-finding algorithms on a broadcast communication model,  $m$ -ary search trees, generalized quicksort, recursive trees, tries, patricia tries, bucket digital search trees, greedy heuristics for matchings on Cayley trees, etc. Further improvements of the approach, yielding a rate to CLT, will also be given.

---

**Svante Janson**

Department of Mathematics, Uppsala University  
svante.janson@math.uu.se

## Quicksort asymptotics

*This is joint work with Jim Fill<sup>1</sup> and is a continuation of his talks in the two preceding seminars in this series.*

I discuss the rate of convergence of the distribution of the (normalized) number of comparisons in Quicksort to its limit distribution. Several different distance measures are considered ( $d_2$ ,  $d_p$ , Kolmogorov-Smirnov, and pointwise difference of Laplace transforms). The results are typically  $O(n^{-1/2})$ , although a minor gap remains for the Kolmogorov-Smirnov distance. This rate is believed to be the true rate of convergence, but rigorous lower bounds are only of the order of  $1/n$ .

---

---

<sup>1</sup>The Johns Hopkins University, jimfill@jhu.edu

Guy Louchard

Département de Mathématiques, Université Libre de Bruxelles  
louchard@ulb.ac.be

## Reflected Brownian Bridge area conditioned on its local time at the origin

*This is a joint work with Philippe Chassaing<sup>1</sup>.*

Throughout this presentation, the standard Brownian motion (BM) will be denoted by  $x(t)$ . Other classical BM are the reflected BM:  $x^+(t) := |x(t)|$ , the Brownian Bridge (BB) on  $[0, 1]$ :  $B(t)$ , the reflected BB on  $[0, 1]$ :  $B^+(t)$ , the Brownian Excursion:  $e(t)$ . The local time of  $x(t)$  at  $a$ , will be denoted by  $t^+(t, a)$  and we define  $Y(b) := \int_0^1 B^+(t) dt$  (area of the reflected BB), conditioned on having a local time at the origin equal to  $b$ . For some reasons that will be clear later on, we propose to call this distribution the "Generalized Airy Distribution".

We are interested in the moments of  $Y(b)$ , i.e.

$$\beta_k(b) := \mathbb{E}[Y(b)^k].$$

We know (see [4], equ(30)) that

$$\int_0^\infty e^{-\alpha t} \mathbb{E}_0 \left[ \exp \left[ - \int_0^t x^+(u) du - \delta t^+(t, 0) \right] \middle| x(t) = 0 \right] \frac{dt}{\sqrt{2\pi t}} = \left[ \delta - \frac{2' A_i'(2'\alpha)}{A_i(2'\alpha)} \right]^{-1} \quad (1)$$

where  $2' := 2^{1/3}$ ,  $A_i$  is the classical Airy function.

In [4], [3] and [1], we were interested in  $X := \int_0^1 e(t) dt$  and we found that this was deeply related to some properties of  $A_i$ . We also obtained a linear recurrence formula for the moments of  $X$ . Motivated by a paper of Janson, [2], we intend here to analyze the corresponding properties of  $Y(b)$ . Starting from (1), this analysis will be directly based on BB and Airy functions properties. The presentation is organized as follows. Part 2 gives the basic results we need in the sequel and relates the moments of  $Y(b)$  to an asymptotic form of the Airy function. Part 3 provides an efficient recurrence formula for the moments. Part 4 is devoted to some asymptotic properties of the  $Y(b)$  density. Part 5 concludes the presentation.

## References

- [1] P. Flajolet and G. Louchard. Analytic variations on the Airy distribution. To appear in *Algorithmica*.
- [2] S. Janson. Asymptotic distributions for the cost of linear probing hashing. <http://www.math.uu.se/~svante/papers/>.
- [3] G. Louchard. The brownian excursion area: a numerical analysis. *Computers and Mathematics with Applications*, 10(6):413–417, 1984.
- [4] G. Louchard. Kac's formula, Levy's local time and brownian excursion. *Journal of Applied Probability*, 21:479–499, 1984.

---

<sup>1</sup>Institut Elie Cartan, INRIA, CNRS and Université Henri Poincaré, [chassain@iecn.u-nancy.fr](mailto:chassain@iecn.u-nancy.fr)

**Hosam M. Mahmoud**

Department of Statistics, George Washington University  
hosam@gwu.edu

## **The Size of Random Bucket Trees via Urn Models**

We find the asymptotic average composition of a class of nonclassic Polya urn models (not necessarily of fixed row sum) by embedding the discrete urn process into a renewal process with rewards. A subclass of the models considered has banded matrix urn schemes and serves as a natural modeling tool for the size of a class of random bucket trees. The class of urns considered extends known results for multicolor urns with constant row sums. The same asymptotic average results are shown to hold in the larger class. This provides an average-case analysis for the size of certain random bucketed multidimensional quad trees and  $k$ -d trees, which are all new results. Some bucket trees have urn schemes with constant row sum, a special case that helps detect phase changes in the limiting distribution of the (normed) size of the tree. For these special cases one can appeal to a more developed urn theory to find a joint limiting distribution of the normed size up to a threshold value of the capacity of a bucket. Once that cut-off point is surpassed, normality ceases to hold. This case appears in paged binary trees (threshold 116),  $m$ -ary search trees (threshold 26), and bucket recursive trees (threshold 26). The asymptotic normality results and the phase change after the threshold in these trees are already known and we only provide alternative proofs via a unified urn models approach.

---

**Jean-François Marckert**

Département de Mathématiques, Université de Versailles Saint-Quentin  
marckert@math.uvsq.fr

## **Non-Crossing trees are almost conditioned Galton-Watson trees**

*This a joint work with Alois Panholzer<sup>1</sup>.*

A non-crossing tree (NC-tree) is a tree drawn on the plane having as vertices a set of points on the boundary of a circle, and whose edges are straight line segments that do not cross. In this talk, we show that NC-trees with size  $n$  are "almost" conditioned Galton-Watson trees. As corollaries, we give the limit depth processes and the limit profile of NC-trees (and so the limit law of the height, of the width, and the total path length).

---

---

<sup>1</sup>Institute of Algebra and Computational Mathematics, Technical University of Vienna,  
apanhol@mail.zserv.tuwien.ac.at

**Conrado Martínez**

Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya  
conrado@lsi.upc.es

## **On the complexity of unranking and ordered generation**

*This is joint work with Xavier Molinero<sup>1</sup>.*

In this work we present generic algorithms for the unranking and ordered generation of labelled combinatorial structures. Unranking means generating the  $i$ -th combinatorial structure of a given size  $n$ , given the rank  $i$  and a specification of the combinatorial class. In the ordered generation, the goal is to generate all combinatorial structures for a given size and specification of the combinatorial class. We also provide a cost algebra for the complexity of these algorithms much in the line of the seminal work of Flajolet et al. on the random generation of labelled combinatorial structures. We show that the unranking algorithms have the same complexity as the corresponding algorithms for random generation and that the ordered generation algorithms will usually be of the same complexity as specifically designed algorithms (where the class to be generated is not a parameter of the algorithm).

---

---

<sup>1</sup>Universitat Politècnica de Catalunya, molinero@lsi.upc.es

Donatella Merlini

Dipartimento di Sistemi e Informatica, Università di Firenze  
merlini@dsi.unifi.it

## Some matrices often arising in combinatorics and in the analysis of algorithms

The concept of Riordan matrices provides a remarkable characterization of many lower triangular arrays that often arise in combinatorics and in the analysis of algorithms. The theory has been introduced in the literature in 1991 by Shapiro et al. [6] and then examined closely from a theoretical and practical point of view in [2, 3, 4, 7]. This study has pointed out that Riordan arrays are a powerful tool in the study of many counting problems.

In this talk we first present the main properties of these matrices referring in particular to some classical example such as Pascal, Catalan, Motzkin and modified Stirling (first and second kind) triangles; we then show some connections between Riordan arrays and both lattice paths and generating trees, a concept, the last one, which is more and more studied in the literature (see, e.g., [1, 5]); finally we describe some examples concerning the applications of the Riordan array concept to the study of some parameters in the job scheduling of a slow device, to the enumeration and the analysis of some data structure such as hybrid trees and to some tiling problems.

## References

- [1] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, and D. Gouyou-Beauchamps. Generating functions of generating trees. *Discrete Mathematics*, to appear.
- [2] D. Merlini. I Riordan Array nell'Analisi degli Algoritmi. Tesi di Dottorato, Università degli Studi di Firenze, 1996.
- [3] D. Merlini, D. G. Rogers, R. Sprugnoli, and M. C. Verri. On some alternative characterizations of Riordan arrays. *Canadian Journal of Mathematics*, 49(2):301–320, 1997.
- [4] D. Merlini, R. Sprugnoli, and M. C. Verri. Waiting patterns for a printer. In *Proceedings of FUN with algorithm'01, The Island of Elba*, 2001.
- [5] D. Merlini and M. C. Verri. Generating trees and proper Riordan Arrays. *Discrete Mathematics*, 218:167–183, 2000.
- [6] L. W. Shapiro, S. Getu, W.-J. Woan, and L. Woodson. The Riordan group. *Discrete Applied Mathematics*, 34:229–239, 1991.
- [7] R. Sprugnoli. Riordan arrays and combinatorial sums. *Discrete Mathematics*, 132:267–290, 1994.

Markus E. Nebel

Institut für Informatik, JWG Universität Frankfurt  
nebel@sads.informatik.uni-frankfurt.de

## RNA Secondary Structures and Their Relation to Binary Trees

The secondary structure of a RNA molecule is of great importance and possesses influence, e.g. on the interaction of tRNA molecules with proteins or on the stabilization of mRNA molecules. The classification of secondary structures by means of their *order* proved useful with respect to numerous applications. In 1978 Waterman, who gave the first precise formal framework for the topic, suggested to determine the number  $a_{n,p}$  of secondary structures of size  $n$  and given order  $p$ . Since then, no satisfactory result has been found. Based on an observation due to Viennot et al. we will derive generating functions for the secondary structures of order  $p$  from generating functions for binary tree structures with Horton-Strahler number  $p$ . These generating functions enable us to compute a precise asymptotic equivalent for  $a_{n,p}$ . Furthermore, we will determine the related number of structures when the number of unpaired bases shows up as an additional parameter. Our approach proves to be general enough to compute the average order of a secondary structure together with all the  $r$ -th moments and to enumerate substructures such as hairpins or bulges in dependence on the order of the secondary structures considered.

---

**Ralph Neininger**

School of Computer Science, McGill University, Montreal  
neiningr@jeff.cs.mcgill.ca

## Multivariate Contraction Method

Well-known contraction arguments for the derivation of limit laws for parameters of random recursive structures and algorithms based on the minimal  $L_2$ -metric (Wasserstein-metric) are extended to recurrences of random vectors of parameters.

This leads besides multivariate limit laws to an easy access to the asymptotic correlation of two parameters. An application is given on the joint behavior of the number of key comparisons and key exchanges of the Quicksort algorithm.

A multivariate point of view may also cover univariate recursions with certain "forbidden" dependences in the recurrence. An example is the Wiener index of a random binary search tree.

Although often asymptotic normality cannot be proven by the minimal  $L_2$ -metric (requiring a change of the metric) some exceptional examples are given, e.g., the number of coin flips in "leader election".

This talk is mainly based on my preprint *On a multivariate contraction method for random recursive structures with applications to Quicksort*.

---

Michel Nguyen-The

Laboratoire d'Informatique de l'X, Ecole Polytechnique  
mnguyen@lix.polytechnique.fr

## Distributions of Valuations on Trees

We consider arithmetical expression trees. The result of the expression is called the *valuation* of the tree. Defining  $X_n$  as the random variable equal to the valuation of a tree of size  $n$  (number of internal nodes), we study the limit distribution of  $X_n$ , for  $n$  going towards  $\infty$ , in particular cases. We will only consider binary trees, and hence binary operators.

We first constrain the leaves (external nodes) to be labelled by integers belonging to a finite set  $I$ , with positive  $P[X_0 = 0]$  and  $P[X_1 = 0]$ , and the internal nodes to be labelled by  $+$  and  $\min$  with probabilities  $\alpha$  and  $\beta = 1 - \alpha$ . For Catalan trees we obtain discrete limits. For binary search trees, the limit depends on  $\alpha$ .

We then consider trees with a distribution of operands having a variance, and operators  $+$  and  $-$  put on internal nodes with probabilities  $\alpha$  and  $\beta = 1 - \alpha$ . For Catalan trees we obtain Gaussian limits (excepted for degenerated cases), while for binary search trees, the limit once again depends on  $\alpha$ .

---

**Daniel Panario**

School of Mathematics and Statistics, Carleton University  
daniel@math.carleton.ca

## **Pairs of coprime smooth polynomials and the Waterloo algorithm**

*This is a joint work with Michael Drmota<sup>1</sup>.*

We focus on the Waterloo variant of the index calculus method for the discrete logarithm problem in finite fields. We provide a rigorous proof for the heuristic arguments for the running time of the Waterloo algorithm. This implies in studying the behavior of pairs of coprime smooth polynomials over finite fields. The proof involves a double saddle point method that is in nature similar to the one of Odlyzko for the rigorous analysis of the basic index calculus.

---

---

<sup>1</sup>TU Wien, Department of Geometry, Michael.Drmota@tuwien.ac.at

**Jordi Petit**

Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya  
jpetit@lsi.upc.es

## **Hamiltonian Cycles in Faulty Random Geometric Networks**

In this talk we shall analyze the Hamiltonian properties of faulty random networks. This consideration is of interest when considering wireless broadcast networks. A random geometric network is a graph whose vertices correspond to points uniformly and independently distributed in the unit square, and whose edges connect any pair of vertices if their distance is below some specified bound. A faulty random geometric network is a random geometric network whose vertices or edges fail at random. Algorithms to find Hamiltonian cycles in faulty random geometric networks are presented.

---

**Boris Pittel**

Department of Mathematics, The Ohio State University  
bgp@math.ohio-state.edu

## Phase transition and finite-size scaling for the integer partitioning problem

*This is a joint work with Jennifer Chayes and Christian Borgs<sup>1</sup>.*

We consider the problem of partitioning  $n$  randomly chosen integers between 1 and  $2^m$  into two subsets such that the discrepancy, the absolute value of the difference of their sums, is minimized. A partition is called perfect if its discrepancy is 0 when the sum of all integers in the original set is even, or 1 when the sum is odd. Introducing  $k = m/n$ , we prove that the problem undergoes a phase transition at  $k = 1$ , in the sense that for  $k < 1$  a.s. there are many perfect partitions, while for  $k > 1$  a.s. there are no perfect partitions. We also determine a scaling window about the transition point  $k(n) = 1 - (\log_2 n + x(n))/(2n)$ , by showing that the probability of a perfect partition tends to 1, 0, or some explicit function  $f(x)$ , depending on whether  $x(n)$  tends to  $-\infty$ ,  $+\infty$ , or  $x$ , respectively. We prove that in the subcritical phase ( $k < 1$ ) the number of perfect partitions is normal in the limit. For the scaling window (critical phase) and the supercritical phase ( $k > 1$ ) we determine the limiting distribution of the (scaled) optimum discrepancy. This appears to be a first case of combinatorial optimization problem for which Derrida's random energy approximation model is proven to be sharp in the limit.

---

---

<sup>1</sup>Theory Group, Microsoft Research

**Nicolas Pouyanne**

Département de Mathématiques, Université de Versailles Saint-Quentin  
pouyanne@math.uvsq.fr

## Permutations admitting an $m$ -th root

The number of permutations of the symmetric group  $S_n$  that admit an  $m$ -th root is asymptotically equivalent to

$$\frac{\pi(m)}{n^{1 - \phi(m)/m}}$$

where  $\phi$  is the Euler function and  $\pi(m)$  an explicit constant.

---

**Helmut Prodinger**

School of Mathematics, University of the Witwatersrand  
helmut@gauss.cam.wits.ac.za

## **Randomized search of fixed length binary words - or - a trie model of two russians**

The traditional model for tries is to flip a coin whenever a decision has to be made. The new russian model chooses a word of length  $n$  (over a two-letter alphabet) for each of the  $m$  data at random and uses its bits. In the limit  $m \rightarrow \infty$ , the classical model resurfaces.

In a joint effort with W. Szpankowski<sup>1</sup> it was possible to compute the average search costs exactly, under several different boundary conditions. The Patricia version and the size of tries can also be successfully handled.

The treatment of digital search trees (under the russian probability model) seems to be a challenge.

---

---

<sup>1</sup>Department of Computer Science, Purdue University, [spa@cs.purdue.edu](mailto:spa@cs.purdue.edu)

**Mireille Regnier**

Projet Algo, INRIA Rocquencourt

Mireille.Regnier@inria.fr

## **Large Deviations on Words**

Word counting satisfies a Large Deviation Principle. The existence of a generating function allows the derivation of an accurate estimate of the tail distribution. Notably, the parameters in Bahadur & Rao Theorem are explicitly computed. Additionally, these results allow for a computation of the distribution of a word conditioned by the overrepresentation of an other word. This methodology applies to assess the statistical significance of exceptional words overrepresentation in computational biology. Related algorithmic and complexity issues are discussed and compared to previous results.

---

**Ludger Rueschendorf**

Institut für Mathematische Stochastik, Universität Freiburg  
ruschen@stochastik.uni-freiburg.de

## A general limit theorem for recursive algorithms

*This talk is based on some joint work with Ralph Neininger<sup>1</sup>. It is a continuation of his recent paper on the multivariate contraction method.*

Our main result gives a limit theorem for divide and conquer type recursions under some general assumptions on the asymptotic behaviour of the first two moments. Using several different metrics this allows to vary the necessary assumptions on the recursion and to analyze a great variety of examples, to obtain e.g. also local type of limit results and in particular to obtain normal limits. The method also allows to establish negative results as in the recent paper of Hwang which shows that for  $m$  greater than 26 no normalization can make the space of  $m$ -ary search trees converge.

---

---

<sup>1</sup>McGill University, [neinigr@jeff.cs.mcgill.ca](mailto:neinigr@jeff.cs.mcgill.ca)

Gilles Schaeffer  
CNRS - LORIA  
Gilles.Schaeffer@loria.fr

## Random planar maps and alternating knots and links

*This is a joint work with Sebastien Kunz-jacques<sup>1</sup>.*

In 1998, Sundberg and Thistlethwaite obtained in a remarkable paper the growth rate of the number of prime alternating knots and links. These are the first classical objects of knot theory for which enumerative results are available.

In order to obtain the precise asymptotic behavior of this number of prime alternating links, we will journey in the realm of analytic combinatorics, and look for properties of random planar maps. The relation between planar maps (embeddings of graphs in the plane) and knots simply arises from the representation of knots by planar diagrams with an under-overcrossing structure on each vertex.

In the review of some almost sure facts about random planar maps (probabilistic properties and conjectures), we shall pick a few ingredients for our enumerative problem. In particular the map-Airy law discussed last year in the analysis of a random sampling algorithm (Banderier, Flajolet, Schaeffer and Soria) will surface here again.

As pointed out by Zinn-Justin and Zuber, planar diagrams representing knots and links can be also viewed as configurations of a toy model on a random lattice. Techniques of mathematical physics then allow to make conjectures extending Sundberg and Thistlethwaite's result. Indeed physicists have already studied how their classical models (Ising, percolation, spanning trees, etc.) are perturbed by replacing the usual regular lattice by a random planar map (this is called "coupling the model with  $2d$  quantum gravity"...). From the combinatorial point of view, these approaches raise a number of intriguing questions.

---

---

<sup>1</sup>Laboratoire d'Informatique (LIX), cole Polytechnique, [kunzjacq@lix.polytechnique.fr](mailto:kunzjacq@lix.polytechnique.fr)

Hadas Shachnai

Department of Computer Science of Haifa  
hadas@cs.technion.ac.il

## Efficient Reorganization of Binary Search Trees

*This is joint work with Micha Hofri<sup>1</sup>.*

We consider the problem of maintaining a binary search tree BST that minimizes the average access cost needed to satisfy randomly generated requests. We analyze scenarios in which the accesses are generated according to a vector of fixed probabilities which is *unknown*.

We devise policies for modifying the tree structure dynamically, using rotations of accessed records. The aim is to produce good approximations of the optimal structure of the tree, while keeping the number of rotations as small as possible. The heuristics that we propose achieve a close approximation to the optimal BST, with lower organization costs than any previously studied.

We introduce the MOVE\_ONCE rule. The average access cost to the tree under this rule is shown to equal the value achieved by the common rule *Move to the Root* (MTR). The advantage of MOVE\_ONCE over MTR and similar rules is that it relocates each of the items in the tree at most once. We show that the total expected cost of modifying the tree by the MOVE\_ONCE rule is bounded from above by  $2(n+1)H_n - 4n$  rotations (in a tree with  $n$  records), where  $H_n$  is the  $n$ th harmonic number. Extensive experiments show that this value is an over-estimate, and in fact the number of rotations is linear for all the access probability vectors we tested. An approximate analysis is shown to match the experimental results, producing the expected number  $n(\pi^2/3 - 2) - 2 \ln n + 0.1354$ .

Next we combine the MOVE\_ONCE

MTR rule with reference counters, one per record, that provide estimates of the reference probabilities. We show that, for any  $\delta, \alpha > 0$  and sufficiently large  $n$ , the resulting reorganization rule achieves a cost that approaches the optimum up to an absolute difference of  $\delta$  with probability higher than  $1 - \alpha$ , within a number of accesses that is proportional to  $n(\lg n)^2/(\alpha\delta^2)$ .

---

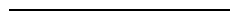
<sup>1</sup>Department of computer science, Worcester polytechnic Institute, [hofri@wpi.edu](mailto:hofri@wpi.edu)

**Renzo Sprugnoli**

Dipartimento di Sistemi e Informatica, Università di Firenze  
sprugnoli@dsi.unifi.it

## Histograms and Fountains

Recently, the concept of a  $p$ -histogram has been introduced to perform a combinatorial study of the queue of a slow device (e.g., a printer) attached to a computer. The concept is equivalent to that of "staircase polyominoes", and precisely a  $p$ -histogram is a sequence of columns (each column composed by cells of the same dimension  $1 \times 1$ ) in which the height of the  $(j + 1)$ th column is at most  $k + p$ , if  $k$  is the height of column  $j$ ; furthermore, the first column has height not greater than  $p$ . In the past,  $p$ -histograms have been studied by number of columns, and in that respect they have strong connections with  $p + 1$ -ary tree. Here we wish to study them by area, which seems to be a more difficult problem. A well-known result concerns the case  $p = 1$  and is due to Odlyzko and Wilf. We use a different approach to count  $p$ -histograms with a given area. This approach is related to the infinite triangles of histograms counted by the height of their last column. Since these histograms represent the queue stories when the total waiting time of users has been fixed, their study can give useful information in dealing with this sort of slow devices.



## Wojciech Szpankowski

Department of Computer Science, Purdue University  
spa@cs.purdue.edu

### New and Old Problems in Pattern Occurrences

Let  $H$  and  $T$  be sequences generated over a finite alphabet. We call  $H$  the pattern and  $T$  the text. The basic questions we are interested in are:

- (i) how many times  $H$  occurs in  $T$ ;
- (ii) how long one has to wait until  $H$  occurs in  $T$ .

We consider the pattern matching problem in various settings, namely:

- (1) In the string matching problem the pattern  $H$  is considered to be a string, while in the subsequence matching problem the pattern  $H$  is a substring.
- (2) The pattern  $H$  can occur exactly or approximately.
- (3) In most cases we assume that the text is generated by a random source and the pattern is given. But in the repetitive pattern matching problem the pattern is a part of the random text.
- (4) Finally, one may assume that the text  $T$  is generated among all possible strings. In the restricted pattern matching problem the text satisfies some additional constraints (e.g., in a  $(d, k)$  sequence there is no runs of 0's shorter than  $d$  and longer than  $k$ ).

In this talk we review analyses of the above mentioned pattern matching problems. We show how to compute moments, limiting distributions and large deviations for the number of pattern occurrences and the waiting time (i.e., the first occurrence of the pattern). Throughout we use combinatorial (e.g., formal calculus) and analytic methods (e.g., generating function, singularity analysis) of analysis of algorithms.

Brigitte Vallée  
GREYC, Université de Caen  
brigitte@info.unicaen.fr

## Hidden pattern statistics

*This is a joint work with P. Flajolet<sup>1</sup>, Y. Guivarch<sup>2</sup> and W. Szpankowski<sup>3</sup>.*

We consider the sequence comparison problem, also known as *hidden pattern (word)* problem, where one searches for a given subsequence in a text (rather than a string understood as a sequence of consecutive symbols). A characteristic parameter is the number of occurrences of a given pattern  $w$  of length  $m$  as a subsequence in a random text of length  $n$  generated by a memoryless source. Spacings between letters of the pattern may either be constrained or not in order to define valid occurrences. We determine the mean and the variance of the number of occurrences, and establish a Gaussian limit law. These results are obtained via combinatorics on words, formal languages techniques, and methods of analytic combinatorics based on generating functions and convergence of moments. The motivation to study this problem comes from an attempt at finding a reliable threshold for intrusion detections, from textual data processing applications, and from molecular biology.

---

---

<sup>1</sup>INRIA Rocquencourt, [Philippe.Flajolet@inria.fr](mailto:Philippe.Flajolet@inria.fr)

<sup>2</sup>IRMAR, Université de Rennes I, [Yves.Guivarch@univ-rennes1.fr](mailto:Yves.Guivarch@univ-rennes1.fr)

<sup>3</sup>Department of Computer Science, Purdue University, [spa@cs.purdue.edu](mailto:spa@cs.purdue.edu)

Alfredo Viola

Instituto de Computacion Facultad de Ingenieria Universidad de la Republica, Montevideo  
viola@fing.edu.uy

## Analysis of Rabin's Irreducibility Test for Polynomials over Finite Fields

*This is a joint work with Daniel Panario<sup>1</sup>, Boris Pittel<sup>2</sup> and Bruce Richmond<sup>3</sup>.*

We give a precise average-case analysis of Rabin's algorithm for testing the irreducibility of polynomials over finite fields. The main technical contribution of this work is the study of the probability that a random polynomial of degree  $n$  contains an irreducible factor of degree dividing several maximal divisors of the degree  $n$ . We then study the expected value and the variance of the number of operations performed by the algorithm and some of its variants. We present an exact analysis when  $n$  is a prime or a product of two primes and an asymptotic analysis for the general case. We also determine the ordering of prime divisors of  $n$  that minimizes the leading factor.

In this talk we will present and motivate the problem, explain the most important technical aspects of our analysis, and show how our analysis can be generalized to other algorithms that deal with similar divisor conditions.

---

---

<sup>1</sup>School of Mathematics and Statistics, Carleton University, [daniel@math.carleton.ca](mailto:daniel@math.carleton.ca)

<sup>2</sup>Ohio State University, [bgp@math.ohio-state.edu](mailto:bgp@math.ohio-state.edu)

<sup>3</sup>University of Waterloo, [lbrichmond@uwaterloo.ca](mailto:lbrichmond@uwaterloo.ca)

**Andreas Weiermann**

Institut für Mathematische Logik, der Westfälischen-Wilhelms Universität Münster  
weierma@math.uni-muenster.de

## **Analytic combinatorics and rapidly growing functions**

Rapidly growing functions can be used as a scale for measuring the complexity of certain combinatorial principles like: Friedman's miniaturization of Kruskal's theorem, Friedman's miniaturization of the well-foundedness of the nested multiset ordering, the termination of the hydra and Goodstein processes and Paris' and Harrington's extension of the finite Ramsey theorem. Surprisingly, results from analytical combinatorics (e.g. Otter 1949, Yamashita 1979) can be used for classifying the complexities of these principles in terms of sub-linear bounding functions. We give an informal survey on this material and we address some related open problems.