

*Project-Team Algo**Algorithms**Rocquencourt*

THEME 2B

The logo features the word "Activity" in a white serif font, with a large, light grey, stylized letter "A" to its left. A horizontal line passes through the middle of the "A" and the word "Activity". Below this, the word "Report" is written in a white serif font, with a large, light grey, stylized letter "R" to its left.

2003

Table of contents

1. Team	1
2. Overall Objectives	1
3. Scientific Foundations	2
3.1. Analysis of Algorithms	2
3.2. Computer Algebra	2
3.3. Algorithms on Sequences	3
4. Application Domains	3
5. Software	3
6. New Results	4
6.1. Analysis of algorithms	4
6.2. Computer Algebra	5
6.3. Algorithms on sequences	6
7. Contracts and Grants with Industry	7
7.1. Industrial Contracts	7
8. Other Grants and Activities	8
8.1. National Actions	8
8.2. Actions funded by the European commission	8
8.3. Bilateral International Relations	8
9. Dissemination	8
9.1. Animation	8
9.2. Teaching	9
9.3. Participation in conferences, seminars, invitations	9
9.4. Foreign Visitors	10
10. Bibliography	10

1. Team

Head of project-team

Bruno Salvy [DR]

Vice-head of project team

Philippe Flajolet [DR]

Administrative assistant

Virginie Collette [TR]

Research scientists (Inria)

Frédéric Chyzak [CR]

Mireille Régnier [DR]

Research scientists (partners)

Cyril Banderier [CR CNRS, University of Paris Nord]

Philippe Dumas [professeur, Cl. Prépa. lycée Jean-Baptiste Say]

Pierre Nicodème [CR CNRS, École polytechnique]

Brigitte Vallée [DR CNRS, University of Caen]

Junior technical staff

Édouard Dolley [Till September 30]

Post-doctoral fellows

Hà Lê [Starting September]

Markus Vöge [Starting September]

José Luis Martins [Starting October]

Ph. D. students

Alin Bostan [École polytechnique, joint supervision]

Marianne Durand [ENS Paris]

Julien Fayolle [University of Paris VI]

Frédéric Giroire [University of Paris VI, starting October]

Ludovic Meunier [École polytechnique]

Vincent Puyhaubert [ENS Cachan]

Mathias Vandenbogaert [University of Bordeaux]

Stagiaire

Frédéric Giroire [University of Paris VI, from April to September]

2. Overall Objectives

The primal objective of the project, inherited from the former century, is the field of *analysis of algorithms*. By this is meant a precise quantification of complexity issues associated to the most fundamental algorithms and data structures of computer science. Departing from traditional approaches that, somewhat artificially, place emphasis on worst-case scenarios, the project focusses on average-case and probabilistic analyses, aiming as often as possible at realistic data models. As such, our research is inspired by the pioneering works of Knuth.

The need to analyse, dimension, and finely optimize algorithms requires an in-depth study of random discrete structures, like words, trees, graphs, and permutations, to name a few. Indeed, a vast majority of the most important algorithms in practice either “make bets” on the likely shape of input data or even base themselves of random choices. In this area we are developing a novel approach based on recent theories of combinatorial analysis together with the view that discrete models connect nicely with complex-analytic and asymptotic methods. The resulting theory has been called—“*Analytic combinatorics*”. Applications of it have been or are currently being worked out in such diverse areas as communication protocols, multidimensional

search, data structures for fast retrieval on external storage, data mining applications, the analysis of genomic sequences, and data compression, for instance.

The analytic-combinatorial approach to the basic processes of computer science is very systematic. It appeared early in the history of the project that its development would greatly benefit from the existence of symbolic manipulation systems and computer algebra. This connection has given rise to an original research programme that we are currently carrying out. Some of the directions pursued include automating the manipulation of combinatorial models (counting, generating function equations, random generation), the development of “automatic asymptotics”, and the development of a unified view of the theory of special functions. In particular, the project has developed the Maple library ALGOLIB, that addresses several of these issues.

3. Scientific Foundations

3.1. Analysis of Algorithms

Key words: *analysis of algorithms, combinatorial analysis, analytic combinatorics, random discrete structures, index tree, hashing methods, asymptotic enumeration, limit law.*

While we know the laws of basic physics and while probabilists have been setting up a coherent theory of stochastic processes for about half a century, the “laws of combinatorics”, in the sense of the laws governing random structured configurations of large sizes, are much less understood. Accordingly, our knowledge in the latter area is still very much fragmentary. Some of the difficulties arise from the large variety of models that tend to arise in real-life applications—the world of computer scientists and algorithmic designers is really an artificial world, much more “free” than its physical counterpart. Some of us have then engaged in the long haul project of trying to offer a unified perspective in this area. The approach of analytic combinatorics has evolved from there.

Analytic combinatorics leads to discovering randomness phenomena that are “universal” (a term actually borrowed from statistical physics) across seemingly different applications. For instance, it is found that similar laws govern the behaviour of prime factors in integers, of irreducible factors in polynomials, of cycles in permutations, and of components in mappings of a finite set. Once detected, such phenomena can then be exploited by specific algorithms that factor integers (a problem relevant to public-key cryptography), decompose polynomials (this is needed in computer algebra systems), reorganize tables in place (this is obvious interest in the manipulation of various data sets), and use collisions to estimate the cardinality of massive data ensembles. The underlying technology bases itself on generating functions, which exactly describe discrete models, as well as an interpretation of these generating functions as analytic transformations of the complex plane. Singularities together with the associated perturbative theory then deliver a number of very precise estimates regarding important characteristics of random discrete structures. The process can be largely made formal and accessible to computer algebra (see below) and it may be adapted to the broad area of analysis of algorithms.

3.2. Computer Algebra

Key words: *asymptotic scales, random generation, special functions, polynomial elimination, Gröbner bases.*

Computer algebra at large aims at making effective large portions of mathematics, paying due attention to complexity issues. For reasons mentioned above, our project specifically investigates the way mathematical objects originating in complex analysis can be dealt with in an algorithmic way by computer algebra systems. Our main contributions in this area concern the automation of asymptotic analysis and the handling of special functions. The mathematical foundations of our algorithms are deeply rooted in differential algebra (Hardy fields for asymptotic expansions and Ore algebras for special functions).

Over the years, in order to automate the average-case analysis of larger and ever larger classes of algorithms, we have developed algorithms and implementations for the following problems: the specification of formally

specified combinatorial structures; the corresponding problems of enumeration and random generation; the automatic construction of asymptotic scales which is necessary for extracting the singular behaviour of generating functions; the automatic computation of asymptotic expansions in such scales; the automatic computation of asymptotic expansions satisfied by coefficients of generating series. An *Encyclopedia of Combinatorial Structures*, available on the web, gathers roughly one thousand structures for which generating series, recurrences, and asymptotic behaviour have been determined automatically using our libraries.

An important principle of computer algebra is that it is often easier to operate with equations defining a mathematical object implicitly rather than trying to obtain a “closed-form” expression of it. The class of linear differential and difference equations is particularly important in view of the large variety of functions and sequences they capture. In this area, we have developed the highly successful GFUN package (jointly with P. Zimmermann, from the Spaces project) dealing with the univariate case. In the multivariate case, we have developed the underlying theory based on Gröbner bases in Ore algebra, and an implementation in the MGFUN package. The algorithmic advances of the past few years have made it possible to start the implementation of an *Encyclopedia of Special Functions*, providing various information concerning classical functions (of wide use throughout sciences), including Bessel functions, Airy functions, The corresponding information is all automatically generated.

3.3. Algorithms on Sequences

Key words: *combinatorics on words, sequences, pattern matching, genome.*

The goal of our research on sequences is the design of new algorithms and the computation of their average case complexity or the derivation of combinatorial results on words and their implementation in statistical software. Possible applications are data compression and genomic sequences. A new area arises in the context of genomic sequences, where biologically significant motifs are extracted. This subject combines searching algorithms of potential signals, the candidates, and computations of statistical significance. For each candidate, the choice criterion is its underrepresentation or overrepresentation. Due to the large number of potential candidates, the speed and the numerical precision of the computation are crucial.

From a methodological point of view, we exhibit several renewal processes, and the limiting law is usually a Gaussian law. Here, the tail distributions are necessary, as one needs to evaluate the overrepresentation, or the underrepresentation, of a motif. The combinatorial properties of words allow, for this class of problems, an effective computation of formulae valid in the central domain and in the tails. Asymptotic analysis yields an exact expression of the rate function, in the sense of large deviation theory. Simultaneously, we define for each problems some characteristic languages in order to bound the computational complexity in the Markovian case.

4. Application Domains

Our work on combinatorial structures applies to modelling and studying complex discrete systems and communication networks. The envisioned applications of the analysis of algorithms are methods for fast access to structured data, fast algorithms in computer algebra and a statistical treatment of biological sequences.

Our areas of research in computer algebra are: combinatorial structures, special functions and sequences, and asymptotic analysis. Our results on special functions lead to algorithms and programs for the automatic treatment of special functions from classical analysis and mathematical physics. In the long term, our work on asymptotic analysis should lead to a bridge between computer algebra and numerical analysis: numerical computations are robust away from singularities and could be complemented by automatically generated code in sensitive areas.

5. Software

The Algolib library is a set of Maple routines that have been developed in the project for more than 10 years. Several parts of it have been incorporated in the standard library of Maple, but the most up-to-date version

is always available for free from our web pages. (The diffusion list for these updates contains more than 200 subscribers). This library provides: tools for combinatorial structures (the `comstruct` package), this includes enumeration, random or exhaustive generation, generating functions for a large class of attribute grammars; tools for linear difference and differential equations (the `gfun` package), which have received a very positive review in *Computing Reviews* and has been incorporated in N. Sloane's `superseeker` at Bell Labs; tools for systems of multivariate linear operators (the `Mgfun` package), including Gröbner bases in Ore algebras, that also treat commutative polynomials and are now the standard way to solve polynomial systems in Maple (although the user does not notice it); `Mgfun` has also been chosen at Risc (Linz) as the basis for their package `Desing`.

We also provide access to our work to scientists who are not using Maple or any other computer algebra system in the form of automatically generated encyclopedia available on the web. The Encyclopedia of Combinatorial Structures thus contains more than 1000 combinatorial structures for which generating series, enumeration sequences, recurrences and asymptotic behaviour have been computed automatically. The Encyclopedia of Special Functions gathers around 40 special functions for which power series, asymptotic expansions, graphs, ...have been generated automatically, starting from a linear differential equation and its initial conditions.

6. New Results

6.1. Analysis of algorithms

Participants: Marianne Durand, Philippe Flajolet, Frédéric Giroire, Vincent Puyhaubert, Mireille Régnier, Bruno Salvy, Brigitte Vallée.

There have been in 2003 two main areas of activity. First, the general theory of *analytic combinatorics*, which serves locally as a basis to a modern vision of the average-case and probabilistic analysis of algorithms, has made progress with a 133 page synthesis report by Flajolet and Sedgewick [20]: it develops basic complex asymptotic methods from first elements of combinatorial theory. Second a number of algorithms of either theoretical or practical interest have been analysed. Furthermore, in the case of a fundamental problem of data mining, a new "universal" algorithm has been designed by Flajolet and Durand [14].

The algorithms analysed are of the following types.

- *Basic sorting and searching algorithms.* Marianne Durand has for the first time provided a complete analysis [8] of Bentley's version of the quicksort algorithm: this is precisely the algorithm used in several versions of the Unix system. Related results on a heuristic model of balanced search trees appear in a joint work [9] of hers with S. Taylor (USA). Finally, in a collaboration with H. Brönnimann (New York) and Frédéric Cazals (Project PRISM, INRIA Sophia), she has developed and analysed precisely a new data structure, the (randomized) "jump-list", which elegantly combines sequential access and logarithmic search costs [13].
- *Data compression algorithms.* Suffix trees are largely used as a data structure for representing texts in the realm of data compression (like in the `gzip` utility) and computational biology. In line with his master's thesis, Julien Fayolle has worked out the average-case behaviour of various parameters of the suffix tree, (like size or external path length) under a memoryless source, thereby proposing an alternative to earlier approaches by Jacquet (HIPERCOM Project) and Szpankowski (Purdue). He is currently extending these results to the broader model of dynamical sources introduced recently by B. Vallée.
- *Random generation and simulation.* With P. Duchon (LaBRI), G. Louchard (Brussels), and G. Schaeffer (LIX), P. Flajolet has developed a brand new approach [7] to the fast generation of complex structured configurations. The framework is inspired by Boltzmann models of statistical physics. The resulting algorithms are often linear (or quasi-linear) in computation time; this has made it possible to routinely generate objects of sizes near 100,000 whereas only sizes of the order of hundreds were known to be attainable by previous methods.

- *Hard combinatorial problems.* Recent years have seen a surge of interest in the probabilistic analysis of instances of hard combinatorial optimization problems. Such questions are especially meaningful in endeavours aimed at overcoming complexity barriers. With D. Gardy (Versailles), B. Chauvin (Versailles), and B. Gittenberger (T.U. Wien), P. Flajolet has shown that the complexity of a Boolean function is somewhat tied to the frequency with which it appears amongst all Boolean computation trees [6]. Vincent Puyhaubert examines similar NP -complete problems, like integer partitioning and the satisfiability of random Boolean formula. On the latter question, he has proposed a new approach based on “urns-and-bins” models that are familiar from probability theory and combinatorial mathematics. This yields in a transparent manner unified proofs for some of the most significant upper bounds known to the satisfiability threshold of random and-or clauses; this problem is itself related to constraint satisfaction in logic programming.

Work has been ongoing regarding the emerging classification of combinatorial processes that are relevant to analysis of algorithms. An instance is a popular survey by P. Chassaing (Nancy) and P. Flajolet [5] concerning the process related to linear probing hashing, that is, to fast access in hashing tables using no extra-storage. It is piquant to note that the occurrence of Brownian excursion area, of some coalescence processes as well as statistics on random trees all have a bearing on the way hashing tables tend to degenerate as they become filled. A new avenue to urn models has been opened in 2003 when P. Flajolet, with J. Gabarro and H. Pekari (Barcelona), have shown for the first time the possibility of developing a purely analytic model of urn processes of the Pólya type [23]. Theoretically, this reveals a classification of certain urn models based on the notion of genus and it leads to significant large deviation estimates, as well as to stable laws or to models exactly solvable in terms of elliptic functions in particular cases. V. Puyhaubert is completing the classification of 2×2 balanced urn models and he has already discovered extension of the framework to 3×3 balanced urns, provided they are of triangular type. Such urn models can additionally describe classical and generalized coupon collector problems as well as balanced data structures of the B -tree type.

Finally, a major new discovery of the year is the *LogLog-Counting* algorithm of M. Durand and P. Flajolet [14]. This new algorithm permits us to estimate the cardinality (understood as the number of *distinct* records) in a huge file using a single pass and only about 2 kilobytes of auxiliary memory for an accuracy of about 1%. This algorithm naturally applies to the gathering of a large number of simultaneous statistics on large “texts”, which may equally well be natural language corpuses in data-mining or router traces in networking. A finely tuned version of the algorithm appears to be totally free of nonlinearities, so that it is unbiased for cardinalities ranging from 1 to 10^9 (say). The algorithm is validated by a thorough mathematical analysis that combines several techniques developed in the project (e.g., generating functions, Mellin transforms, saddle-point methods). At the same time, it has been tested extensively on various sets of natural data; examples include 200 millions digits of π , extensive http server traces, Shakespeare’s complete works, the *Mahābhārata* Indian epic and the *Rg Veda*, to name a few. Frédéric Giroire is starting a Ph. D. Thesis (Fall 2003) on an important class of problems in data mining corresponding to the extraction of quantitative information from very large amounts of data using only a very small memory.

6.2. Computer Algebra

Participants: Alin Bostan, Frédéric Chyzak, Hà Lê, Ludovic Meunier, Marni Mishna, Bruno Salvy.

L. Meunier develops an *Encyclopedia of Special Functions*, that is generated *automatically*. The underlying algorithms and implementations are based in particular on works by B. Salvy and P. Zimmerman (the GFUN package), and by F. Chyzak (the MGFUN package). This encyclopedia gathers identities, formulae and graphs that are computed automatically from the data of a linear differential equation and initial conditions. All the production process being automated, the difficult and expensive step of checking each formula individually is suppressed. Available on the web (<http://algo.inria.fr/esf>), this encyclopedia also plays the rôle of a showcase for part of the packages developed in our project. Its design has been presented at the international computer algebra conference ISSAC’03 [16].

For several years, F. Chyzak and P. Paule (RISC, University of Linz, Austria) have been collaborating on the writing of a chapter on computer algebra methods for special functions, in the framework of the project Digital Library of Mathematical Functions (DLMF) of the National Institute of Standards and Technology (NIST). This ambitious project aims at providing a new edition for the “Handbook of Mathematical Functions,” an authoritative handbook since 1962 and probably the most cited work in the history of scientific publications. The chapter is mainly concerned with those algorithms that are at the heart of the GFUN and MGFUN packages. A draft was finalized this year, and the project is entering a stage of validation by external experts. The book will be available both in printed version (roughly 1,000 pages) and under electronic format (a CD and a web site, see <http://dlmf.nist.gov/>).

Yet, a longer-term goal of NIST will be to make full use of advanced communications channels and automated calculation tools, so as to present not only static data, but also dynamical pieces of information, produced on demand, such as function graphs, numerical tables, and, even, tables of mathematical identities and symbolic transformations. The authoritative nature of the existing handbook and its orientation towards applications within sciences, statistics, engineering and calculations will be preserved; but its utility value will be largely extended, far beyond the traditional limitations of printed documents, making DFLM a vehicle that revolutionizes practice and diffusion of applied mathematics in general. It goes without saying that Meunier’s ESF shares this goal of more interactivity, and that each project should benefit from the experience gained by the other.

A recent follow up to F. Chyzak and B. Salvy’s work is the application of methods originally developed for special functions to deal with symmetric functions in algebraic combinatorics. Last year’s collaboration with Marni Mishna resulted in an algorithm for the computation of scalar products between symmetric series, making possible the enumeration of classes of graphs given by regularity constraints. In 2003, the collaboration on this topics has continued, leading to new symmetric functions identities with a representation-theoretic interpretation and to asymptotic results on the enumeration of regular graphs.

The problem of extracting special subseries of a formal power series described by a defining linear differential equation—extracting the even terms, the odd terms, etc—appears naturally in various areas of mathematics, such as the computer algebraic study of multisummable divergent series. A natural algorithm for this task uses ALGO’s usual tools: Hadamard products of differentially finite functions, which can be computed by the GFUN package; but other equally natural candidates have been proposed, based on Galois theory. Together with M. Barkatou and M. Loday-Richaud, F. Chyzak worked on comparing various methods and optimizations, and on understanding the variations in the output representations [12].

H. Le continues his research started during his PhD study at the Symbolic Computation Group of the University of Waterloo, Canada. He finalized a paper on telescoping series in the context of symbolic summation [3]. He also worked on several optimizations of Zeilberger’s summation algorithm, and on normal forms for rational functions to be used in the context of symbolic summation and integration.

For several years, B. Salvy has been working jointly with the STIX laboratory of the *École polytechnique*. This work applies recent algorithmic progress on straight-line programs in order to produce efficient algorithms and implementations for geometrical problems. Now, the aim is to extend these methods based on geometric resolution to the non-commutative context necessary for the application to special functions. This has been part of the PhD thesis of A. Bostan, jointly supervised by M. Giusti and B. Salvy, and defended this year [2]. As a first step, an extension of the commutative toolbox has yielded efficient algorithms for eliminations involving *two* algebraic numbers [21]. The asymptotic complexity of these algorithms is optimal, up to logarithmic factors. As a second step, these ideas extend to the general case of polynomial systems. They lead to faster algorithms for minimal polynomials in some quotient algebras, that lead to fast computation of parameterization of the solutions [4].

6.3. Algorithms on sequences

Participants: Edouard Dolley, Philippe Flajolet, Pierre Nicodème, Mireille Régnier, Bruno Salvy, Mathias Vandenbogaert.

Analytic combinatorics has allowed the team to solve numerous word or sequence problems: (i) one or several motifs, possibly infinite families, regular expressions, palindromes, ... (ii) exact or degenerate motifs; (iii) various probability models (Bernoulli, Markov, dynamic sources, ...). Such analyses allow us to construct "toolkits" that allow to distinguish a significant signal from the noise, in several domains in computer science (text data, security systems, genomic data, ...).

Our study of the distribution relies on the definition and manipulation of specific languages the generating functions of which satisfy algebraic equations systems. This leads to explicit formulae, for the mean, variance, and the tail distribution. It turns out that their actual computation simplifies when the set of words has a specific structure. In a collaboration with J. Clément (Marne-la-Vallée University), M. Régnier proposed a general definition of the approximation that is consistent with the biological constraints on the so-called regulatory signals, and designed an efficient algorithm to compute the moments under a Markovian model. M. Régnier and A. Denise (Orsay University) studied the tail distributions [17] for word occurrences. They derived an exact expression of the rate function and an asymptotic expansion of the probabilities. They also derived closed formulae for conditional expectations, that allow to extract a weak signal hidden by a stronger signal. The computational complexity turns out to be polynomial. An extension to couples of words allowed to deal with the important case of double strand counting. An application on plants data sets, with M. Lescot (Marseille University) shows the accuracy of the method, that overcomes less sophisticated software such as RSA-tools [15].

M. Vandebogaert studies palindrome avoidance in bacterial genomes in a collaboration with M. S. Gelfand and V. Makeev (NII-Genetika). He extends a previous study undertaken by E. M. Panina, A. A. Mironov et M. S. Gelfand to approximate palindromes. The key idea is that under-represented palindromes are likely to be selected against if they are binding sites for the restriction enzymes of the restriction-modification system of that genome. M. Vandebogaert, in a PhD thesis to be defended soon, points out the noise introduced when the errors are uncontrolled and limits them to the ones allowed by the IUPAC code. His analyses, using the combinatorial results described above, have revealed some interesting examples of possible horizontal transfer events of RM systems during the evolution. Preliminary results have been presented at BGRs'02 and MCCMB'03 [11][29].

These results are implemented in C or Maple procedures. E. Dolley realised a web site, to allow their use on-line. Notably, in a collaboration supported by the grant GenomAI, some of these procedures have already been imported in RSA-tools, developed at the *Université Libre de Bruxelles* by J. van Helden for the detection of regulatory motifs.

A collaboration with F. Tahi and S. Engelen (Évry university) describes an extension of a previous algorithm to predict secondary structures, that allows to take into account pseudo-knots [18].

We collaborate on this subject with other INRIA projects. Algorithmic and combinatorial approaches of ADAGE (G. Kucherov) are complementary to our combinatorial and probabilistic approach (see [27]), for instance on hidden words (see Flajolet's work) or tandem repeats. These results find applications in the SMILE software developed by L. Marsan and M. F. Sagot (Helix).

7. Contracts and Grants with Industry

7.1. Industrial Contracts

The Algorithms Project and Waterloo Maple Inc. (WMI) have developed a collaboration based on reciprocal interests. It is obviously interesting for the company to integrate functionalities at the forefront of the current research in computer algebra. Reciprocally, this integration makes our programs and our research visible to a very wide audience.

Numerous exchanges have thus taken place between the project and the company over the years. After more than 3 years within the project, J. Carette has been for several years Product Development Director at WMI, before going back to the academic world. Similarly, E. Murray, who worked for two years in the project developing the combstruct package is now working at WMI.

Thanks to all this activity, the company WMI considers Inria as a special partner and grants it a free license for all of its research units. Moreover, a cooperation agreement has been signed between WMI and ALGO in 2001. In particular, one of the objectives is to replace all the routines dealing with asymptotic and series expansions in Maple by implementation of new algorithms dealing with very general classes of asymptotic scales.

8. Other Grants and Activities

8.1. National Actions

M. Régnier animates the project *Algorithmique et statistique des séquences* at the IMPG (*Informatique, Mathématique et Physique pour le Génome*).

Aléa is a national working group dedicated to the analysis of algorithms and random combinatorial structures. It is a meeting place for mathematicians and computer scientists working in the area of discrete models. It is currently supported by CNRS (GDR A.L.P.) and is animated by Philippe Flajolet. In 2003, the yearly meeting has gathered 70 participants from about 15 different research laboratories throughout France.

8.2. Actions funded by the European commission

The Algo project is, for a period of three years (2000-2003), one of the components of the project ESPRIT "Long Term Research" ALCOM-FT. This project gathers ten leading groups in the field of algorithmic research in Europe. The objective is to find new algorithmic concepts and identify transverse key algorithms of many applications. Four directions of work have been identified: (i) Massive data sets; (ii) Communication systems; (iii) Optimisation in production and planning; (iv) Methodological and experimental algorithmic research. Work of our project is mainly in the axes (ii) and (iv).

8.3. Bilateral International Relations

Mireille Régnier is responsible for the grant GENOMAL in the Bioinformatics program, that involves the École polytechnique and the universities of Marseille and Bordeaux. Together with Alain Denise (Orsay) she was responsible of the group "Algorithms, Combinatorics and Statistics in Genomic Sequences". She is the French scientific head of a Liapunov project.

The ALGO project, the Vienna University of Technology, and RISC (University of Linz) are partners in an Amadeus project (bilateral action between France and Austria), financed by Égide. PhD student Stefan Gerhold (RISC, Austria) spent two weeks with the ALGO project. A goal in his thesis will be to elaborate on the symbolic summation algorithms designed by F. Chyzak and B. Salvy. In this framework, Frédéric Chyzak and Michael Drmota (Vienna) have initiated a joint collaboration on the distribution of the number of occurrences of patterns in random trees; Marianne Durand and Michael Drmota have continued the work started last year in the same project on the extraction of coefficients in bivariate generating series by double saddle-point contours.

9. Dissemination

9.1. Animation

The ALGO project runs a biweekly seminar. Several partner teams in the grand Paris area attend on a regular basis. Proceedings are edited and published yearly [1].

Philippe Flajolet has served as member of the program committee of ICALP'03, the pan-European conference on theoretical computer science, held this year in Malaga (Spain). He continues to serve as Chair of the Steering Committee of the series of *Seminars on Analysis of Algorithms*, which this year took place in San Miniato (Italy) with about 70 participants. He is also responsible for the French *Aléa* group (under the auspices

of GDR ALP) dedicated to the study of random structures and algorithms, which in 2003 organized a meeting of some 65 participants from mathematics and computer science. He has served as member of the evaluation committee of ACI-NUM, a concerted action of the French Ministry of Education dedicated to the new interfaces of mathematical sciences. He is an editor of the journal *Random Structures and Algorithms* and an honorary editor of *Theoretical Computer Science*. He also serves as an editor of Cambridge University Press' prestigious series "*Encyclopedia of Mathematics and its Applications*". He has served in thesis committee of Baert (Amiens, reporter), Gillet (Nancy, reporter), Bostan (École polytechnique), and a few others. In 2003, he has served on the Committee on the Teaching of Mathematics commissioned by the Ministry of Education. He is a member of the Recruiting Committee for computer science at École polytechnique. With G. Huet, he has given a colloquium talk at "Centre d'Alembert" in Orsay on "Mathematics and Computer Science, the case of algorithmics". On November 18, 2003, Philippe Flajolet has been elected as a Member (Fellow) of the French Academy of Sciences (in the Mechanical Sciences section).

Mireille Régnier co-organized the 1st Moscow Conference on Computational Molecular Biology, July, 22nd-25th, that was supported by ERCIM and INRIA. She organized with Pierre Nepomiaschty the stand of the Liapunov Institute at France Tech exhibition in Moscow (October, 7-10), and the colloquium for the 10-th anniversary of Liapunov Institute. She was a member of the PhD committee of J. Bourdon (Caen).

Bruno Salvy has been a member of the steering committee of ISSAC, which is the premier international conference in computer algebra, where his term ended this year. He is a member of the recruitment committee of the *Université des Sciences et Technologies de Lille* (in computer science) and of the University of La Rochelle (in mathematics). He is also member of the editorial board of the *Journal of Symbolic Computation*. For next year, he is a member of the program committee of the conference ISSAC'04 and of that of the first French-Canada congress in Mathematical Sciences, that will be held in Toulouse in 2004. He has been a member of the PhD committees of Marni Mishna (*Université du Québec à Montréal*) and Alin Bostan (*École polytechnique*). He maintains and animates the French mailing list devoted to the computer algebra system Maple.

9.2. Teaching

Frédéric Chyzak, teaches several computer science courses as a *chargé d'enseignement à temps incomplet* at École polytechnique, including one in computer algebra.

Marianne Durand is teaching at the university of Versailles–Saint-Quentin, where she gives the plenary course in computer science for the first years (first semester).

Philippe Flajolet has only taught a 10 hour fifth-year course in the Joint Master Course in Algorithms of the wider Paris area, this due to a recent reorganization of the curriculum by authorities.

Frédéric Giroire is teaching a "System" class in License, a "Networking" class and a "Performances Analysis" class in the Second year of IUP at the Paris VII university.

Vincent Puyhaubert teaches the Java programming language at University of Versailles to first year DEUG students.

Mireille Régnier teaches a 10 hours post-graduate course on "Combinatorics and Genome" at Évry, and a 25 hours graduate course at the École Centrale de Paris, on the "Mathematical Problems and Algorithms in Genomics".

Mathias Vandenbogaert occasionally teaches topics related to bioinformatics at the École Centrale de Paris.

9.3. Participation in conferences, seminars, invitations

Frédéric Chyzak presented results on computer algebra manipulations of special functions and combinatorial sequences at the University of Limoges, and the work on the calculation of scalar products of symmetric functions during the *Journées nationales de Calcul formel* at CIRM and at the University of Lyon.

Marianne Durand presented a poster in San Miniato (Italy). She gave a talk at the ESA03 conference in Budapest, at the Algo seminar, at the algorithmic seminar of the university of Caen, and also at a conference for young researchers in Marne-La-Vallée. She gave two talks at the Aléa conference in Marseille. She visited

Michael Drmota of the University of Vienna (Austria), and Alfredo Viola in Montevideo (Uruguay). She gave a talk in both places.

Julien Fayolle gave a talk at the Algorithms Project's Seminar on "Suffix trees and simple sources".

Philippe Flajolet has presented talks on "analytic urns" at the following places: 50th "Séminaire Lotharingien de Combinatoire", Probability Seminar at the Free University of Brussels, 9th International Seminar on Analysis of Algorithms in San Miniato, "Journées Arbres" at Versailles, Algorithms Project at Rocquencourt, as well as at the Universities of Dijon (mathematics department) and Caen (algorithms seminar). He has been invited at a physics seminar of École Normale Supérieure where he gave a lecture on *Singular Combinatorics*. He has given a talk at the computer science seminar of École polytechnique on Boltzmann models and random combinatorial generation. He has been invited to give a Colloquium talk at INRIA Sophia on *Analysis of Algorithms*, the video being available on the web site of INRIA Sophia. He has also been invited to give a lecture within the framework of "Leçons de mathématiques d'aujourd'hui" at Bordeaux, with title "The digital search tree".

Hà Lê gave a talk at the seminar in computer algebra at Limoges University. He gave a talk at the Algorithms Project's Seminar.

Ludovic Meunier gave talks about the Encyclopedia of Special Functions at the *Journées Nationales du Calcul Formel* (JNCF, Luminy), at the *École des Jeunes Chercheurs* (EJC, Marne la Vallée), at the LORIA project (INRIA, Nancy) and at the ISSAC 2003 conference (Philadelphia).

Vincent Puyhaubert gave talks at the Algorithms Project's Seminar, Aléa workshop and AofA seminar. On these three occasions, he presented his work on threshold phenomena in satisfiability problems. He also presented some new results on analytic urns at Algorithms Project's Seminar.

Mireille Régnier presented her results at BIBE'03 (Washington, DC), at "Journées de Statistiques" in Lyon, MCCMB'03 (Moscow) and was an invited speaker at the Drug and Discovery Bioinformatics conference in Mainz (Germany).

Bruno Salvy has been invited to spend a week at the *Université du Québec à Montréal*, where he gave a talk on the analysis of connected graphs. He has given a presentation on *Algorithms for D-finite functions* at the French workshop on computer algebra, in Luminy. He has given a course on this same subject, with a combinatorial bias, at the yearly meeting of the *Aléa* work group.

Mathias Vandenbogaert gave a survey talk about his research, in the LaBRI, University of Bordeaux I. He presented a poster in Berlin and another one in Moscow.

Markus Vöge has been invited to give a talk at Saint Gallen, Switzerland.

9.4. Foreign Visitors

A large number of our visitors have gives talks at the seminar of the project. This year, we received: Stefan Gerhold, University of Linz (Amadeus); Anatoly Vershik, St.-Petersburg Mathematical Society; George Labahn, University of Waterloo; Wojtek Szpankowski, Purdue University, jointly with the Hipercom project; John Shackell, University of Kent; Marni Mishna, University of Quebec at Montreal; Conrado Martinez, Universitat Politècnica de Catalunya; Cristoph Richard and Thomas Prellberg, University of Greifswald; Will James, University of Melbourne; Vsevolod Makeev, NIIGenetika, Moscow (Liapunov).

10. Bibliography

Books and Monographs

- [1] *Algorithms Seminar, 2001–2002*. F. CHYZAK, editor, Institut National de Recherche en Informatique et en Automatique, series Research Report, volume 5003, November, 2003, <http://www.inria.fr/rrrt/rr-5003.html>.

Doctoral dissertations and “Habilitation” theses

- [2] A. BOSTAN. *Algorithmique efficace pour des opérations de base en calcul formel*. Ph. D. Thesis, École polytechnique, 2003.

Articles in referred journals and book chapters

- [3] S. ABRAMOV, J. CARETTE, K. GEDDES, H. LE. *Telescoping in the Context of Symbolic Summation in Maple*. in « Journal of Symbolic Computation », To appear.
- [4] A. BOSTAN, B. SALVY, É. SCHOST. *Fast Algorithms for Zero-Dimensional Polynomial Systems Using Duality*. in « Applicable Algebra in Engineering, Communication and Computing », number 4, volume 14, 2003, pages 239–272.
- [5] P. CHASSAING, P. FLAJOLET. *Hachage, Arbres, Chemins, et Graphes*. in « Gazette des Mathématiciens », 2003, pages 29–49.
- [6] B. CHAUVIN, P. FLAJOLET, D. GARDY, B. GITTENBERGER. *And/Or Trees Revisited*. in « Combinatorics, Probability and Computing », To appear, 27 pages.
- [7] P. DUCHON, P. FLAJOLET, G. LOUCHARD, G. SCHAEFFER. *Boltzmann Samplers for the Random Generation of Combinatorial Structures*. in « Combinatorics, Probability and Computing », To appear, 44 pages.
- [8] M. DURAND. *Asymptotic Analysis of an Optimized Quicksort Algorithm*. in « Information Processing Letters », volume 85, 2003, pages 73–77.
- [9] M. DURAND, S. TAYLOR. *Emerging Behavior as Binary Search Trees Are Symmetrically Updated*. in « Theoretical Computer Science », volume 297, 2003, pages 425–445.
- [10] F. TAHI, M. GOUY, M. RÉGNIER. *Automatic RNA secondary structure prediction with a comparative approach*. in « Computers & Chemistry », number 5, volume 26, 2002, pages 521–530.
- [11] M. VANDENBOGAERT, V. MAKEEV. *Analysis of bacterial RM-systems through genome-scale analysis and related taxonomic issues..* in « In Silico Biology », number 3, volume 12, 2003, pages 127–143.

Publications in Conferences and Workshops

- [12] M. BARKATOU, F. CHYZAK, M. LODAY-RICHAUD. *Remarques algorithmiques liées au rang d’un opérateur différentiel linéaire*. in « From Combinatorics to Dynamical Systems (Journées de Calcul Formel, Strasbourg, March 22-23, 2002) », series IRMA Lectures in Mathematics and Theoretical Physics, number 3, de Gruyter, F. FAUVET, C. MITSCHI, editors, pages 87-129, 2003.
- [13] H. BRÖNNIMANN, F. CAZALS, M. DURAND. *Randomized Jumplists: A Jump-and-Walk Dictionary Data-Structure*. in « Proceedings of the STACS’03 Conference, Berlin, February 2003 », series Lecture Notes in Computer Science, volume 2607, Springer Verlag, pages 283–294, 2003.
- [14] M. DURAND, P. FLAJOLET. *Loglog Counting of Large Cardinalities*. in « Annual European Symposium

on Algorithms (ESA03) », series Lecture Notes in Computer Science, volume 2832, G. DI BATTISTA, U. ZWICK, editors, pages 605–617, September, 2003.

- [15] M. LESCOT, M. RÉGNIER. *Motif statistics on plants datasets*. in « MCCMB'03 », 2003, Proc. Moscow Conference on Computational Molecular Biology, to appear in *Biophysica*.
- [16] L. MEUNIER, B. SALVY. *ESF: An Automatically Generated Encyclopedia of Special Functions*. in « Symbolic and Algebraic Computation », ACM Press, J. R. SENDRA, editor, pages 199–205, 2003, Proceedings ISSAC'03, Philadelphia, August 2003..
- [17] M. RÉGNIER, A. DENISE. *Statistiques Extrêmes sur les Mots*. in « XXXV-èmes Journées de Statistique », volume 2, pages 799–802, 2003, In Proceedings XXXV-èmes Journées de Statistique, Lyon.
- [18] F. TAHI, S. ENGELEN, M. RÉGNIER. *A fast algorithm for RNA secondary structure prediction including pseudoknots approach*. in « BIBE'03 », IEEE Computer Society, pages 11–17, 2003.

Internal Reports

- [19] M. BARDET, J.-C. FAUGÈRE, B. SALVY. *Complexity of Gröbner basis computation for Semi-regular Overdetermined sequences over $GF(2)$ with solutions in $GF(2)$* . Research Report, number 5049, Inria, December, 2003, <http://www.inria.fr/rrrt/rr-5049.html>, 19 pages.
- [20] P. FLAJOLET, R. SEDGEWICK. *Analytic Combinatorics: Basic Complex Asymptotics*. Research Report, number ALCOM-FT-TR-300, ALCOM Project, November, 2003, 133 pages. Chapters 4 and 5 of the book project *Analytic Combinatorics*.

Miscellaneous

- [21] A. BOSTAN, P. FLAJOLET, B. SALVY, É. SCHOST. *Fast Computation with Two Algebraic Numbers*. September, 2003, 41 pages. Submitted to the *Journal of Symbolic Computation*..
- [22] J. A. FILL, P. FLAJOLET, N. KAPUR. *Singularity Analysis, Hadamard Products, and Tree Recurrences*. Preprint, June, 2003, 47 pages. Submitted to *Journal of Computational and Applied Mathematics*..
- [23] P. FLAJOLET, J. GABARRÓ, H. PEKARI. *Analytic Urns*. March, 2003, 52 pages. Submitted to *Annals of Probability*.
- [24] P. FLAJOLET, W. SZPANKOWSKI, B. VALLÉE. *Hidden Word Statistics*. Preprint, 2003, 34 pages; submitted to *Journal of the ACM*..
- [25] F. GIROIRE. *Etude de Problèmes Combinatoires liés à l'Analyse du Génome : Séquençage et Polymorphisme*. 2003, Mémoire de DEA, Université Paris VI..
- [26] M. RÉGNIER, A. DENISE. *Rare Events on Random Strings*. 2003, Submitted.
- [27] M. RÉGNIER. *Mathematical Tools for Regulatory Signals Extraction*. 2003, to appear in Kluwer series.

-
- [28] M. VANDENBOGAERT. *Bacterial RM-systems revisited using degenerate pattern consensus*. Poster, 2003, presented at Recomb 2003, The Seventh Annual International Conference on Research in Computational Molecular Biology, Berlin, Germany..
- [29] M. VANDENBOGAERT. *Bacterial RMS revisited: avoided words reflect homology through protein sequence similarity*. Poster, 2003, presented at Moscovian Conference in Computational Molecular Biology (MCCMB 2003) in Moscow, Russia, July 22-25, 2003..