INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Algo

Algorithms

Rocquencourt

THEME SYM

Table of contents

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)
Alin Bostan [CR, starting October]
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]

Post-doctoral fellows
Hà Lê [Until September]
José Luis Martins [Until October]
Markus Vöge [Until April]

Ph. D. students
Marianne Durand [ENS Paris, Ph.D. defense in March]
Julien Fayolle [University of Paris VI]
Éric Fusy [Corps des Telecoms]
Frédéric Giroire [University of Paris VI]
Ludovic Meunier [École polytechnique]
Vincent Puyhaubert [ENS Cachan]
Mathias Vandenbogaert [University of Bordeaux, Ph.D. defense in March]

Students intern
Claudia Coupra [University of Paris XIII, from October to December]
Pierre Guilleminot [École polytechnique, from April to June]
Bao-Truc Lê [University of Évry, from March to June]
Madhur Tulsiani [Internship Program, from May to July]

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 scenarii, 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

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

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

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

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

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

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 `combstruct` 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, under developement by L. Meunier, gathers around 40 special functions for which identities, power series, asymptotic expansions, graphs, ...have been generated automatically, starting from a linear differential equation and its initial conditions. The underlying algorithms and implementations are those of GFUN and MGFUN. 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.

6. New Results

6.1. Analysis of algorithms

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

There have been in 2004 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 synthesis report of over 600 pages by Flajolet and Sedgewick [2]: it develops basic complex asymptotic methods from first elements of combinatorial theory and results in a precise quantification of a great many properties of random discrete structures. (See also the book [1] edited by Drmota, Flajolet, Gardy, and Gittenberger for a wide range of approaches to random combinatorics and algorithms.) Second a number of algorithms of varied theoretical and practical interest have been conceived and/or analysed.

The algorithms designed and analysed are of the following types.

Basic sorting and searching algorithms. Marianne Durand has defended her PhD thesis [3] in April 2004 at the École Polytechnique. In it, she presents a complete combinatorial analysis of hashing strategies based on random probing in the context of a paged memory structure. This part of her work is based on joint work with A. Viola (Montevideo) and it combines urn models, random allocations, and an original variant of the Laplace method of asymptotic analysis.

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 in [22] 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. With M. Ward (Purdue), he also obtained results on the average depth of insertion in suffix trees under the Markovian model.

Compact encoding of mesh graphs. Eric Fusy has obtained, in collaboration with D. Poulalhon (University of Paris 7) and G. Schaeffer (LIX), a very efficient algorithm [24] for encoding the combinatorial structure of a polygonal mesh, a problem which has become prominent in the theory of mesh compression. It relies on an elegant bijection between binary trees and a family of planar graphs. The coding is optimal in the sense that it matches the entropy of the number of meshes of the corresponding size. In addition, Eric Fusy has analysed the algorithm, proved its correctness and established that its time-complexity is linear. Finally, a part of the algorithm consists in finding the minimal Schnyder woods of a 3-connected planar graph, which has many useful applications to graph drawing.
Random generation and simulation. With P. Duchon (LaBRI), G. Louchard (Brussels), and G. Schaeffer (LIX), P. Flajolet has developed a brand new approach 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. A detailed study of 49 pages has recently appeared [9]. E. Fusy is currently investigating an extension of this framework dedicated to the difficult problem of generating random planar graphs uniformly at random.
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 [8]. Vincent Puyhaubert examines similar NP-complete problems, like integer partitioning and the satisfiability of random Boolean formula (in CNF forms, that is, as conjunctions of clauses). On the latter question, he has proposed a new approach [14] 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.
Arithmetical sequences and cryptanalysis. Symmetric cryptographic primitives like block ciphers are typically constructed from a small set of simple building blocks like bitwise exclusive-or and addition modulo a power of 2. Differential cryptanalysis is an attack of ciphers based on the propagation of differences in functions. Philippe Dumas in collaboration with H. Lipmaa and and J. Wallén (Helsinki University of Technology) has developed an original approach [27] relating some of these questions to the theory of regular and automatic integer sequences. (These otherwise surface in several areas of theoretical computer science, including deterministic divide-and-conquer algorithms, formal languages, and the theory of sequential circuits by Vuillemin.) This study neatly points to some of the complexity inherent in the asymptotic behaviour of such integer sequences, while eventually paving the way to a complete classification theorem.

Work has been ongoing regarding the emerging classification of combinatorial processes that are relevant to analysis of algorithms. Phenomena involving Gaussian laws amongst discrete structures are by now fairly well understood, either through the classical theory of staochastic processes or within the framework of analytic combinatorics. Work conducted within the group has revealed next the importance of coalescences and confluences that are conducive to Airy phenomena. In this context, Flajolet, Salvy, and Schaeffer [11] have provided a new analytic approach to the analysis of connectivity in random graphs. On a related register, for a great many probabilistic models encountered in discrete mathematics, singularities provide extremely precise and valuable information. The article [10] written by P. Flajolet in collaboration with J. Fill and N. Kapur (Johns Hopkins University) studies some classical tree models (binary search trees, Catalan trees, union-find trees) but not so standard toll functions. Functions amenable to singularity analysis are shown to be closed under Hadamard product (i.e., the termwise product of series). A valuable consequence is the possibility of classifying the solution to several basic recurrences of the probabilistic divide-and-conquer type whose central role in the design of efficient algorithms is well recognized.

A new avenue to urn models has been opened 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 [12]. 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. In his PhD thesis, V. Puyhaubert completes the classification of 2×2 balanced urn models while discovering extensions 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, balanced data structures of the B-tree type, as well as a simple model of conflicts (Flajolet and Puyhaubert, in preparation).

Finally, a major new discovery of the period 2003-2004 is the LogLog-Counting algorithm of M. Durand and P. Flajolet (see M. Durand's thesis [3] for the latest developments). 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 is 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 Mahabharata Indian epic and the Rg Veda, to name a few. Frédéric Giroire works on a Ph. D. Thesis relative to 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 and he has already obtained promising results concerning estimates based on minima. A summary of this area of research is given by Flajolet in [23].

6.2. Computer Algebra

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

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 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 work already most cited 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 last year after interaction of NIST, and the project is now in a stage of validation by external experts. After a one-year delay, the result of the project is expected to be published next year. 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 is utilitary value will be largely extended, far beyond the traditional limitations of printed documents, making DFLM a vehicle that revolutionarizes practice and diffusion of applied mathematics in general. It goes without saying that the 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 developped for special functions to deal with symmetric functions in algebraic combinatorics. Over the last two years, a collaboration with Marni Mishna resulted in algorithms for the computation of scalar products between symmetric series, making possible the enumeration of classes of graphs given by regularity constraints and

leading to new symmetric functions identities with a representation-theoretic interpretation and to asymptotic results on the enumeration of regular graphs. Collaboration on this topics is continuing.

Another follow up of methods for special functions is an application to linear control systems [21]. A collaboration of F. Chyzak with A. Quadrat (Café Project, INRIA-Sophia-Antipolis) and D. Robertz (University of Aachen, Germany) has shown that elimination methods for non-commutative polynomials designed in the project permit to make methods developped by A. Quadrat for the recognition of properties of linear control systems effective. The spectrum of applications includes ODEs, PDEs, multidimensional discrete systems, differential time-delay systems, repetitive systems, multidimensional convolutional codes, etc. A package, OreModule, http://wwwb.math.rwth-aachen.de/OreModules/, has been developped, based on F. Chyzak's Maple implementation of Gröbner tools.

One of the crucial algorithms implemented in Mgfun is Chyzak's generalization of a classical algorithm by Zeilberger. Yet, there are still issues related to the efficiency of Zeilberger's algorithm that are not fully understood. Ha Le worked on this topic this year. He finalized a paper on telescoping series in the context of symbolic summation [5]. He also worked on several optimizations of Zeilberger's summation algorithm [18], and on normal forms for rational functions to be used in the context of symbolic summation and integration [26][25].

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. In particular, this year, one of the steps of this method has been improved in the case of a variety which is not irreducible. The net result is an algorithm of a much wider applicability: it factors bivariate polynomial with a complexity which is for the first time less than quadratic [20].

Now, the aim is to extend these methods based on geometric resolution to the non-commutative context necessary for the application to special functions. As a first step, it is necessary to obtain low complexity algorithms for algorithms based on evaluation and interpolation in the commutative case. In this vein, Alin Bostan pursued his research on the design of fast algorithms for basic operations in computer algebra. In [6], sharp complexity estimates are given for the problems of multipoint evaluation and interpolation with respect to various polynomial bases and for special families of evaluation points, such as geometric and arithmetic progressions. Moreover, it is shown in [7] that the questions of multipoint polynomial evaluation and interpolation are computationally equivalent in a strong sense.

6.3. Algorithms on sequences

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

Analytic combinatorics 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 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. A survey on the use of generating functions in a related area, e.g. sequence alignments and secondary structures, written by M. Régnier and F. Tahi (Evry University), is presented in [16].

In a recent work [15], M. Régnier and A. Denise (Orsay University) studied the tail distributions for word occurrences. The combinatorial structure of the words allowed for the derivation of an exact expression of the rate function and an asymptotic expansion of the probabilities. A first application is the extraction of a weak signal hidden by a stronger signal. This is made possible by the conditional results derived in [15]. A second application is the assessment of the significance of clustered signals. This work is currently studied with E. Panina (NII Genetika). The formulas have been implemented for the Markov model. The problem reduces to the solution of a polynomial equation; therefore, the computational complexity is low. Indeed, the work of

Maxime Kormilitsine for his master thesis has shown that these results are more accurate and precise than other computations (R'MES and SPA) that have an exponential cost. These results allowed M. Régnier and M. Vandenbogaert to participate to an international contest organized by M. Tompa (Washington University) between statistical softwares for InSilico prediction of regulatory signals. This contest is discussed in [17]. 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 the popular, but less sophisticated, software RSA-tools [13].

M. Vandenbogaert defended a Phd thesis [4] that deals with the assessment of the functional importance of oligonucleotides in genomic sequences. The main biological result is the assessment of horizontal transferring events for the Restriction-Modification System (RMS) in micro-organisms The key idea is to point out a correlation between the under-representation of palindroms and phylogeny. Indeed, palindroms are potential binding sites for the restriction enzymes of the RMS; hence, they are likely to be selected against. His combinatorial results on word underrepresention, and its programs, allowed to validate the palindrome avoidance hypothesis.

M. Vandenbogaert pointed out the noise introduced when the errors are uncontrolled and limited them to the ones allowed by the IUPAC code. In a collaboration with J. Clément (Marne-la-Vallée University), M. Vandenbogaert and M. Régnier proposed a general definition of approximation that is consistent with the biological constraints on the so-called regulatory signals. Combinatorial formulas that allow for computing the waiting time for approximate words, either in the usual case or under this restriction, are given in [4]. J. Clément and M. Régnier are currently working to design an efficient algorithm to compute these formulas. A first application is given for tandem repeats [19]. Tandem repeats are short repetitions that are hotspots for genome recombinations and are also linked to some genetic diseases. Word counting procedures have been implemented in C or Maple. A library of C procedures to be downloaded, or reused by statistical softwares for motif discovery, has been written by M. Tulsiani.

We collaborate on this subject with other INRIA projects. The algorithmic and combinatorial approach of ADAGE (G. Kucherov) is complementary to our combinatorial and probabilistic approach, for instance on hidden words (see Flajolet's work) or tandem repeats. Some of our results find applications in the software SMILE 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 globally animated by Philippe Flajolet. In 2004, the yearly meeting (organized by B. Vallée and A. Akhavi) has gathered in Luminy over 80 participants from about 20 different research laboratories throughout France.

For the period 2003-2006, the Algo Team participates in ACI-NIM a national research programme exploring New Interfaces of Mathematics. In this context, we take part in the ACPA project dedicated to paths and trees, probabilities and algorithms, this jointly with the Universities of Versailles, Bordeaux, and Nancy. In 2004, a project called FLUX and involving the RAP Project at INRIA as well as the University of Montpellier has been funded for a three year period by the national action ACI-MD relative to massive data: our objective is to develop high performance algorithms for the quantitative analysis of massive data flows an important problem in the monitoring of high speed computer networks.

Frédéric Chyzak and Bruno Salvy were among the organizers of a workshop in Toulouse on “Links between Numerical Analysis and Computer Algebra” that gathered 30 participants from various parts of the world.

8.2. Actions funded by the European commission

The ALGO project is one of the components of the project ESPRIT “Long Term Research" ALCOM-FT, which has terminated in 2004. This project gathers ten leading groups in the field of algorithmic research in Europe. The objective is to find new algorithmic concepts and identify key generic algorithms accross 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 has been mainly in the axes (ii) and (iv) and conducted jointly with the Project-team RAP of INRIA-Rocquencourt.

8.3. Bilateral International Relations

Mireille Régnier is the French scientific head of a Liapunov project. This project, jointly with an Armenian team and a Georgian team is supported by the French program ECO-NET.

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.

Julien Fayolle gave talks at the Universities of Dijon, Nantes and Purdue (Illinois, USA). He talked at the “Journées Arbres" in Versailles University, at the AofA Colloquium at UC Berkeley (California, USA) and at the ALÉA meeting in Marseilles. He also presented his results at the *Third colloquium on Computer Science and Mathematics* at the Technical University of Vienna (Austria). All of these talks were centered around the analysis of parameters in suffix tries.

Philippe Flajolet has continued to serve as Chair of the Steering Committee of the series of Seminars on Analysis of Algorithms, which this year took place in Berkeley (USA) with about 75 participants. He is also responsible for the French Aléa group (under the auspices of CNRS and GDR ALP) dedicated to the study of random structures and algorithms, which in 2005 organized a meeting of over 80 participants from mathematics and computer science. He has served as member of the evaluation committee of ACI-NIM, 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, an honorary editor of Theoretical Computer Science, and an honor member of the French association SPECIF. He also serves as an editor of Cambridge University Press' prestigious series "Encyclopedia of Mathematics and its Applications". He is a member of the Recruiting Committee for computer science at École polytechnique. In June 2004, Philippe Flajolet has been officially received as a Member (Fellow) of the French Academy of Sciences (in the Mechanical Sciences section). In December 2004, he has been awarded the Silver Medal of CNRS for his contributions to research in computer science. (Such a distinction is awarded only every second year to a computer scientist in France.) He has been a member of the programme committee of the Third Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities (September 13–17, 2004, Vienna, Austria) and is co-editor of the proceedings, a 550 page volume published by Birkäuser. In 2004, he has served in thesis committees of Michel Nguyen-The and Marianne Durand (both defended at École Polytechnique). He finally serves as member of the College of Reviewers for the Canada Research Chairs Program (mathematics and computer science).

Mireille Régnier M. Régnier was a member of the program committee of RECOMB'04 satellite meeting on Regulation. She organized in Erevan (Armenia) a 10 days school "Combinatorics and Genome". She was a member of the PhD committee of M. Vandenbogaert (Bordeaux).

Bruno Salvy was a member of the program committee of this year's edition of the conference ISSAC, which is the premier international conference in computer algebra. He was also in the program committee of the first French-Canada congress in Mathematical Sciences, that was held in Toulouse in July. 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* and of the *Journal of Algebra* (section Computational Algebra). This year, he has been a member of the PhD committees of Thomas Cluzeau (University of Limoges), Magali Bardet (University Paris 6), Yvan Leborgne (University of Bordeaux), Elie Mosaki (University of Lyon), Pascal Giorgi (ENS Lyon) and a referee for the thesis of Guillaume Chèze (University of Nice).

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. Together with Bruno Salvy, Marc Giusti, François Ollivier and Éric Schost (the latter 3 are at École polytechnique), he teaches a course in computer algebra in the *Master Parisien de Recherche en Informatique* (MPRI).

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

Philippe Flajolet has taught a 15 hour fifth-year course in the Joint Master Course in Computer Science Research (MPRI) of the wider Paris area.

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". She gave a few courses on "Computational Biology" in the MPRI.

9.3. Participation in conferences, seminars, invitations

Frédéric Chyzak has presented joint work in progress with Ph. Dumas, H. Lê, J. Martins, M. Mishna, and B. Salvy (all from ALGO, INRIA) in a talk by the title of "Taming Apparent Singularities via Ore Closure" at an international workshop on algebraic and analytic aspects of (q-)difference equations (Lille, France).

Philippe Flajolet has been the Invited Plenary Speaker at the First Workshop on Analytic Algorithmics and Combinatorics (ANALCO04), New Orleans, January 2004. He has been one of the four invited speakers at ICALP'04 (Turku, Finland, July 2004), which is the major European conference in theoretical computer science in Europe. He has been (together with Don Knuth and Persi Diaconis) one of the keynote speakers at the Tenth Seminar on Analysis of Algorithms AofA'04 (Berkeley, USA, June 2004). He has also given the Opening Keynote Address at the Ninth Asian Computing Science Conference ASIAN'04 (Chiang-Mai, Thailand, December 2004). Philippe Flajolet has been invited to teach three postgraduate courses of 10 to 12 lectures each in Barcelona (April 2004, Polytechnic University of Catalonia, doctoral programme), Berkeley (June 2004, under the auspices of the Mathematical Sciences Research Institute), and Chiang-Mai, Thailand (the ASIAN'04 postconference school). He has otherwise given lectures and seminars at Luminy and Caen.

Frederic Giroire has been invited to spend two weeks in the Inria project Mascotte (Sophia Antipolis). There, he gave a presentation on random counting algorithms, and worked on minimizing tolerant networks for telecommunication satellites.

Vincent Puyhaubert presented his work on "Analytic urns of triangular form" at the Alea workshop (Marseille) and at the AofA conference (MSRI, Berkeley).

Mireille Régnier presented her results at Nantes, Lille, École polytechnique, Hopital René Huguenin (Saint-Cloud) and University of South California. She was invited by INTAS to BGRS'04 for a prospective workshop on the collaboration between EC and NEI in Life Science.

Bruno Salvy gave a presentation on the complexity of Gröbner bases at the "Analysis of Algorithms" conference held at MSRI, Berkeley. He gave a talk on this same subject in a seminar at Marne-la-Vallée and in a workshop "Arbres, Chemins : Probabilités et Algorithmes" in Nancy.

9.4. Foreign Visitors

A large number of our visitors have gives talks at the seminar of the project. This year, we received: Omer Gimenez (Universitat Politècnica de Catalunya, Barcelona, Spain), Hsien-Kuei Hwang (Academia Sinica, Taiwan), Marni Mishna (Simon Fraser U., Canada), Agnes Szanto (North Carolina State U., USA), Mark Ward (Purdue U., USA), Mark Wilson (U. of Auckland, New Zealand).

10. Bibliography

Books and Monographs

[1] M. DRMOTA, P. FLAJOLET, D. GARDY, B. GITTENBERGER (editors). *Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities*, Trends in Mathematics (Mathematics, Computer Science), 554 pages, Birkhäuser Verlag, 2004.

[2] P. FLAJOLET, R. SEDGEWICK. *Analytic Combinatorics*, Chapters I–IX of a book to be published by Cambridge University Press, 609p.+x, available electronically from Philippe Flajolet's home page, November 2004.

Doctoral dissertations and Habilitation theses

[3] M. DURAND. *Combinatoire analytique et algorithmique des ensembles de données*, Ph. D. Thesis, École polytechnique, 2004.

[4] M. VANDENBOGAERT. *Algorithmes et mesures statistiques pour la recherche de signaux fonctionnels dans les zones de régulation*, Ph. D. Thesis, INRIA – LaBRI, 2004.

Articles in referred journals and book chapters

[5] S. A. ABRAMOV, J. J. CARETTE, K. O. GEDDES, H. Q. LE. *Telescoping in the Context of Symbolic Summation in Maple*, in "Journal of Symbolic Computation", vol. 38, nº 4, October 2004, p. 1303–1326.

[6] A. BOSTAN, É. SCHOST. *Polynomial evaluation and interpolation on special sets of points*, in "Journal of Complexity", Festschrift for Arnold Schhönhage, To appear.

[7] A. BOSTAN, É. SCHOST. *On the complexities of multipoint evaluation and interpolation*, in "Theoretical Computer Science", vol. 329, nº 1–3, December 2004, p. 223–235.

[8] B. CHAUVIN, P. FLAJOLET, D. GARDY, B. GITTENBERGER. *And/Or Trees Revisited*, in "Combinatorics, Probability and Computing", Special issue on Analysis of Algorithms, vol. 13, nº 4–5, 2004, p. 501–513.

[9] P. DUCHON, P. FLAJOLET, G. LOUCHARD, G. SCHAEFFER. *Boltzmann Samplers for the Random Generation of Combinatorial Structures*, in "Combinatorics, Probability and Computing", Special issue on Analysis of Algorithms, vol. 13, nº 4–5, 2004, p. 577–625.

[10] J. A. FILL, P. FLAJOLET, N. KAPUR. *Singularity Analysis, Hadamard Products, and Tree Recurrences*, in "Journal of Computational and Applied Mathematics", vol. 174, February 2005, p. 271–313.

[11] P. FLAJOLET, B. SALVY, G. SCHAEFFER. *Airy Phenomena and Analytic Combinatorics of Connected Graphs*, in "The Electronic Journal of Combinatorics", 30 pages, vol. 11, nº 1, May 2004, R34.

[12] P. FLAJOLET, J. GABARRÓ, H. PEKARI. *Analytic Urns*, in "Annals of Probability", 30 pages, To appear.

[13] M. LESCOT, M. RÉGNIER. *Motif statistics on plants datasets*, in "Biophysica", Preliminary version in MCCMB'03, To appear.

[14] V. PUYHAUBERT. *Generating functions and the satisfiability threshold*, in "Discrete Mathematics and Theoretical Computer Science", vol. 6, nº 2, 2004, p. 425–436.

[15] M. RÉGNIER, A. DENISE. *Rare events and Conditional Events on random strings*, in "Discrete Mathematics and Theoretical Computer Science", vol. 6, nº 2, 2004, p. 191–214.

[16] M. RÉGNIER, F. TAHI. *Generating Functions in Computational Biology*, in "Journal of Iranian Statistics", Preliminary version at MABS'97, To appear.

[17] M. TOMPA, N. LI, T. BAILEY, G. CHURCH, B. D. MOOR, E. ESKIN, A. FAVOROV, M. FRITH, Y. FU, J. KENT, V. MAKEEV, A. MIRONOV, W. NOBLE, G. PAVESI, G. PESOLE, M. RÉGNIER, N. SIMONIS, S. SINHA, G. THIJS, J. VAN HELDEN, M. VANDENBOGAERT, Z. WENG, C. WORKMAN, C. YE, Z. ZHU. *An Assessment of Computational Tools for the Discovery of Transcription Factor Binding Sites*, in "Nature Biotechnology", To appear.

Publications in Conferences and Workshops

[18] S. A. ABRAMOV, H. Q. LE. *Utilizing relationships among linear systems generated by Zeilberger's algorithm*, in "Formal Power Series and Algebraic Combinatorics", Proceedings of FPSAC'04, Vancouver, June 2004, 2004, p. 29–38.

[19] V. BOEVA, V. MAKEEV, M. RÉGNIER. *SWAN: searching for highly divergent tandem repeats in DNA sequences and statistical significance*, in "JOBIM'04", In Proceedings JOBIM'04, Montréal, IEEE Computer Society, 2004.

[20] A. BOSTAN, G. LECERF, B. SALVY, É. SCHOST, B. WIEBELT. *Complexity Issues in Bivariate Polynomial Factorization*, in "Symbolic and Algebraic Computation", J. GUTIERREZ (editor)., Proceedings of ISSAC'04, Santander, July 2004, ACM Press, 2004, p. 42–49.

[21] F. CHYZAK, A. QUADRAT, D. ROBERTZ. *OreModules, A symbolic package for the study of multidimensional linear systems*, in "Sixteenth International Symposium on Mathematical Theory of Networks and Systems, Leuven, Belgium", Proceedings MTNS2004, Katholieke Universiteit Leuven, Belgium, July 5–9, 2004, Katholieke Universiteit Leuven, July 2004.

[22] J. FAYOLLE. *An average-case analysis of basic parameters of the suffix tree*, in "Mathematics and Computer Science", M. DRMOTA, P. FLAJOLET, D. GARDY, B. GITTENBERGER (editors)., Proceedings of a colloquium organized by TU Wien, Vienna, Austria, September 2004, Birkhäuser, 2004, p. 217–227.

[23] P. FLAJOLET. *Counting by coin tossings*, in "Proceedings of ASIAN'04 (Ninth Asian Computing Science Conference)", M. MAHER (editor)., Lecture Notes in Computer Science, Text of Opening Keynote Address, vol. 3321, 2004, p. 1–12.

[24] E. FUSY, D. POULALHON, G. SCHAEFFER. *Dissections and trees, with applications to optimal mesh encoding and to random sampling*, in "Proceedings of the 16th Annual ACM-SIAM Symposium On Discrete Mathematics (SODA-05), New York", ACM Press, January 2005, p. 23–25.

[25] K. GEDDES, H. LE, Z. LI. *Differential rational normal forms and a reduction algorithm for hyperexponential functions*, in "Symbolic and Algebraic Computation", J. GUTIERREZ (editor)., Proceedings of ISSAC'04, Santander, July 2004, ACM Press, 2004, p. 183–190.

[26] H. LE, Z. LI. *Differential rational normal forms and representations of hyperexponential functions*, in "Rhine Workshop on Computer Algebra", Proceedings of RWCA'04, Nijmegen, March 2004, 2004, p. 3–12.

[27] H. LIPMAA, J. WALLÉN, P. DUMAS. *Differential Probability of Exclusive-Or*, in "Fast Software Encryption 2004", B. ROY, W. MEIER (editors)., Lecture Notes in Computer Science, Delhi, India, February 5–7, 2004, vol. 3017, Springer-Verlag, 2004, p. 317–331.

[28] M. RÉGNIER. *Mathematical Tools for Regulatory Signals Extraction*, in "Bioinformatics of Genome Regulation and Structure", N. KOLCHANOV, R. HOFESTAEDT (editors)., Preliminary version at BGRS'02, Kluwer Academic Publisher, 2004, p. 61–70.