INRIA

Project-Team Algo

Algorithms

Rocquencourt

THEME SYM

Activity Report

2006

Table of contents

1. Team

Head of project-team
Bruno Salvy [Research Director (DR) Inria]

Vice-head of project team
Philippe Flajolet [Research Director (DR) Inria, HdR]

Administrative assistant
Virginie Collette [(TR) Inria]

Research scientists (Inria)
Alin Bostan [Research Associate (CR) Inria]
Frédéric Chyzak [Research Associate (CR) Inria]
Mireille Régnier [Research Director (DR) Inria, HdR]

Research scientists (partners)
Philippe Dumas [professeur, Cl. Prépa. lycée Jean-Baptiste Say]

Visiting scientists
Micha Hofri [Worcester Polytechnic Institute, Worcester, MA, USA, until June]

Ph. D. students
Julien Fayolle [University of Paris VI, until March]
Éric Fusy [Détachement du Corps des Télécoms]
Frédéric Giroire [University of Paris VI, until September]
Carine Pivoteau [University of Paris VI]

PostDoc
Frédéric Meunier [Détachement du Corps des Ponts et Chaussées, from september]

Students intern
Bruno Besson [University of Paris XI, from March to September]

Technical Staff
André Derrick Balsa [until November]

2. Overall Objectives

2.1. 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 focuses 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 analyze, 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 laws*, *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 of 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*, *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 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 algorithms searching for 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. Software

4.1. 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 into the standard library of Maple, but the most up-to-date version is always available for free from our web pages. 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 have been incorporated in N. Sloane's superseeker at Bell Labs; tools for systems of multivariate linear operators (the *Mgfun* package), including GrËbner bases in Ore algebras, that also treat commutative polynomials and have been the standard way to solve polynomial systems in Maple for a long period (although the user would 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 encyclopedias 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 behavior have been computed automatically. It gets more than 16,000 hits per month. The Encyclopedia of Special Functions 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 (fhttp://algo.inria.fr/esf/), this encyclopedia also plays the role of a showcase for part of the packages developed in our project. It gets 27,000 hits per month.

A new package, *MultiSeries* has been developed recently. It implements so-called multi-series, that are series in general asymptotic scales, each of whose coefficient is itself potentially a new series. This makes it possible to handle in a transparent and dynamic way the problems of finding the proper asymptotic scale for an expansion and of dealing with the indefinite cancellation problem. This package is designed in such a way that it can take the place of the existing *series*, *asympt* and *limit* Maple functions, in a totally transparent manner.

Tandem repeats are short repetitions that are hotspots for genome recombinations and are also related to some genetic diseases. *TandemSWAN* searches for degenerate tandem repeats without insertions and deletions, but with a high substitution rate. It is based on calculation of the repeat statistical significance and identifies the length of the repeated unit and the number of repetitions. It allows for the identification of weak clustered sites, that are needed for the analysis of several important regulatory systems such as nitrate/nitrite switch. It is written in C language, using some C++ features.

P-DCFold implements in Java a heuristic algorithm for the prediction of RNA secondary structures including all kinds of pseudoknots. It is based on the comparative approach, and its input is a small set of RNA sequences. It has been applied to *tmRNA* and *RnaseP* sequences.

5. New Results

5.1. Analysis of algorithms

Participants: André Derrick Balsa, Frédéric Chyzak, Philippe Flajolet, Éric Fusy, Frédéric Giroire, Frédéric Meunier, Mireille Régnier, Bruno Salvy.

The reference book on analytic combinatorics [1] has almost reached completion.

Combinatorial enumeration can often be reduced to setting up equations that translate a given combinatorial-probabilistic model. It has been shown in [9] that an important class of urn models (of relevance in particular to data structures and algorithms) can be treated systematically.

The asymptotic number of occurrences of patterns in random trees obeys a normal low [7]. Moreover, the numerical constants involved in the asymptotic combinatorial results can be computed explicitly by algorithms presented in that article. Noticeably, these algorithms base on symbolic algorithms to manipulate trees, reminiscent of the unification algorithm.

The precise probabilistic analysis of pattern occurrences in sequences is another valuable outcome of analytic combinatorics. We note here a paper [12] published in the highly selective *Journal of the ACM*, which predicts the statistically unavoidable subsequences to be observed in a random text, a problem originally motivated by intrusion detection in computer security.

In general, unlabeled graphs are difficult to handle analytically because of their internal symmetries. The asymptotic enumeration of unlabeled outerplanar graphs has been dealt with [31]. This is an important subfamily of planar graphs. In addition, several asymptotic properties have been analyzed, regarding the number of edges, the chromatic number and connectedness. More generally, unlabeled structures can be handled by techniques descending from Pólya's works [21]. The key point is the introduction of an unbiased pointing operator. A striking application is the possibility of sampling unrooted unlabeled structures uniformly at random without using rejection. This makes it possible to sample unrooted trees and unlabeled graphs of very large size. Another random generator, for plane partitions, is given in [20]. These partitions are extensively studied in statistical physics. The generator is strikingly efficient, being sublinear to generate a plane partition under the Boltzmann distribution.

A new efficient algorithm to draw a quadrangulation on a regular grid has been given in [26]. It exploits some constrained orientations and relies on simple face-counting operations.

Tangible progress has been attained in enlarging the class of functions amenable to the method of singularity analysis, (e.g., closure through Hadamard products, with a quaint turn-of-the-1900's mathematical flavour). This has applications to a number of tree data structures [10].

An important data structure for dictionaries, the digital tree, is at the basis of numerous algorithms on words, including compression algorithms. These structures can be analyzed very precisely by methods from analytic combinatorics [2], [24].

Recently, we designed the LOGLOG counting algorithm. With a few thousand bits of memory, this algorithm makes it possible to estimate the number of distinct elements in a stream of data of several tens of gigabytes, with an accuracy of few percent (typically 2% with 2048 bytes). Another algorithm, MINCOUNT has been adapted to maintain running counts in a "sliding window", without degradation of efficiency [28]. In this context, no a priori probabilistic assumption is made on the nature of data—the algorithms are universal. These algorithms have applications in the area of data mining and especially network monitoring. The current state-of-the-art is given in [27], with an algorithm that has been completely analyzed and validated on traffic traces. It proves very efficient and relevant in the context of traffic monitoring.

5.2. Computer Algebra

Participants: Alin Bostan, Frédéric Chyzak, Philippe Flajolet, Bruno Salvy.

A pedagogical introduction to computer algebra is given in [16].

The manipulation of algebraic numbers gives rise to special bivariate resultants. While no quasi-optimal algorithm is known in general for bivariate resultants, it is shown in [5] that composed sums and composed products can be computed fast, using the first power sums of the roots of polynomials as an intermediate data-structure.

The computation of polynomial and rational solutions of linear recurrences with polynomial coefficients is at the heart of many algorithms related to indefinite or definite summation. In [22], we showed how these solutions can be computed efficiently. A key remark is that these solutions are very structured. This allows to avoid the computation of their coefficients and, instead, to represent them by a compact data structure: initial conditions and a recurrence. This idea of change of representation makes it possible to speed-up the definite and indefinite summation in the classical hypergeometric case studied by Zeilberger.

High precision expansions of power series solutions of differential equations are needed in various branches of computational mathematics, from combinatorics, where the desired power series is a generating function, to numerical analysis and computational number theory. In [23], we give fast algorithms for computing many coefficients of power series solutions of systems of differential equations. The new algorithms use a number of arithmetic operations which is quasi-linear in the number of computed terms. Moreover, these algorithms are optimal, in the sense that the cost of resolution is proportional to that of checking the solution.

In [29], we applied the previous result to a problem in elliptic curve cryptography. We introduced an algorithm that computes a special rational morphism between two elliptic curves; its complexity is quasi-linear with respect to the degree of the morphism. The new algorithm allows to speed-up a crucial step of the Schoof-Elkies-Atkin algorithm for determining the cardinality of an elliptic curve over a finite field.

For several years, B. Salvy and A. Bostan have been working jointly with the LIX laboratory of the *École polytechnique*. This work applies recent algorithmic progress on straight-line programs and has produced efficient algorithms and implementations for geometrical problems. Recently, this work has taken a new direction by extending to the numerical universe methods originally designed to deal with multiplicities when searching for symbolic solutions of polynomial systems. The results obtained by B. Salvy, G. Lecerf (University of Versailles Saint-Quentin-en-Yvelines), M. Giusti (École polytechnique) and J.-C. Yakoubsohn (University of Toulouse) are new versions of Newton's algorithm that are quadratically convergent even in the neighborhood of a multiple root or a cluster of roots for polynomial systems [15] under a technical condition of "embedding dimension 1".

5.3. Algorithms on sequences

Participants: Bruno Besson, Philippe Flajolet, Mireille Régnier, Bruno Salvy.

Our goal is to combine analytic combinatorics to establish general mathematical results and algorithms that provide efficient optimized implementations for specific data constraints.

Our group studies combinatorial properties of words in order to improve searching algorithms and extract unexpectedly frequent or rare motifs from random sequences. Our main motivation is to relate such motifs to biological functions on the genome. Ultimately, statistical or probabilistic results are to be integrated into string searching algorithms or software.

Starting from exact counting formulae derived by the tools of analytic combinatorics, we focus on efficient implementations and tight approximation derivations. On the one hand, efficient implementations have been realized for sets of words, that rely on automata theory. Our main application is the study of words, also called sites on the genome, that may be recognized by different proteins, in a competing or synergistic manner. On the other hand, approximations have been derived, with a tight error bound. Recent results with A. Denise (Orsay University) allow to compute the probability of rare events, commonly called *p-value* by bioinformaticians, with a linear algorithm. This is a drastic improvement on the existing exponential algorithm. Tightness of these *p-value* expressions allows to compute conditional probabilities, and extract weak signals hidden by stronger signals. *P-value* computation was implemented in software *ScanSeq* developed at NIIGenetika, with whom we collaborate. As a whole, systematic comparisons of commonly used statistical criteria for exceptional words, and precise statements on their validity domains, have been realized in [17]. In his master's thesis [30], B. Besson studied the mathematical grounds of the so-called Position Specific Scoring Matrix model, PSSM, in favour among biologists. Notably, he established a link between the number of available experimental data used to build the model, and the precision and significance of the model obtained. A special attention has been paid to Tandem Repeats. Tandem repeats are short repetitions that are hotspots for genome recombinations and are also related to some genetic diseases. *TandemSWAN* searches for degenerate tandem repeats without insertions and deletions, but with a high substitution rate. It is based on calculation of the repeat statistical significance and identifies the length of the repeated unit and the number of repetitions. It allows for the identification of weak clustered sites, that are needed for the analysis of several important regulatory systems such as nitrate/nitrite switch. The originality of the software *TandemSWan*, developed jointly in C with NIIGenetika, is the implementation of a procedure to assess the statistical significance of repeats. Such results can be implemented in other software, including *Mreps* developed at Inria Futurs (Lille) by G. Kucherov. In a collaboration with D. Papatsenko (UC Berkeley), *TandemSwan* results led to establish a relationship between tandem repeats and regulatory mechanisms.

RNA secondary structures prediction is a hard problem, as a given sequence has an exponential number of potential foldings. It is even worse when pseudo-knots are taken into account. Finding a structure, e.g., a set of helices, that is common to sequences from phylogenetically closed organisms turned out to be the most

efficient way. This comparative approach is meaningful as sequencing is often realized now for closely related genomes. Our heuristic relies on a few simple probability criteria to chose common helices, and led recently to a variant, *P-DCfold*, that deals with pseudo-knots [18]. It has been applied to *tmRNA* and *RnaseP* sequences.

6. Contracts and Grants with Industry

6.1. 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.

7. Other Grants and Activities

7.1. National Actions

Alea 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 2005, the yearly meeting (organized by C. Lavault) has gathered in Luminy over 80 participants from about 20 different research laboratories throughout France.

For the period 2003–2006, the Algo project 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.

Since last year, 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.

For the period 2006–2009, the Algo project participates in a programme funded by the National Research Agency (ANR) entitled GECKO for "A Geometric Approach to Complexity and its Applications". Four teams are involved: ALGO (coordinator) and teams at the École polytechnique, the Universities of Toulouse and Nice. The project concentrates on three classes of objects: (i) univariate and multivariate polynomials (Newton process, factorization, elimination); (ii) structured matrices (whose coefficients can be polynomials); (iii) linear differential operators (noncommutative elimination, integration). The aim is to improve significantly the resolution of systems of algebraic or linear differential equations that appear in models, by taking geometry into account.

The National Research Agency (ANR) has funded this year a research project entitled SADA, whose goal is to investigate fundamental properties of random discrete structures and algorithms. The project duration is 3 years (Dec. 2005–Dec 2008). It involves five teams: ALGO/RAP from INRIA Rocquencourt, the Universities of Caen, Versailles, and Bordeaux (coordinator), as well as the Laboratory for Computer Science of the École polytechnique (LIX).

7.2. Bilateral International Relations

Mireille Régnier is the French scientific head of a bioinformatics project supported by the French program ECO-NET, that involves three teams from Armenia, Kazakhstan and Russia.

She is a participant of IST INTAS: This is a grant to work on comparative genomics of bacteria. This project is joint with TUM (Munchen), CNR-ITB (Milan), NII-Genetika (Moscow) and Moscow University.

8. Dissemination

8.1. Animation

The ALGO project runs a biweekly seminar devoted to the analysis of algorithms and related topics. Several partner teams in the grand Paris area attend on a regular basis, and also take part in a yearly workshop, Alea. Proceedings are collected and edited.

Alin Bostan has been a member of the program committee of this year's edition of the ISSAC conference (Genoa, Italy), the premier international conference in computer algebra.

Frédéric Chyzak is a co-organizer of the next edition of the French national meeting in computer algebra (JNCF'07), that will gather some 80 participants in Luminy. He is a member of the recruiting committee of Univ. Limoges, in mathematics.

Philippe Flajolet is an editor of the journal Random Structures and Algorithms, an honorary editor of Theoretical Computer Science, and an honour member of the French association SPECIF. He also serves as one of the three editors of Cambridge University Press' prestigious series "Encyclopedia of Mathematics and its Applications".

He serves as Chair of the Steering Committee of the international series of Conferences and Workshops called "Analysis of Algorithms". The yearly edition attracts some 80 specialists of the area. He serves in a similar capacity as founder and chair of the French Working Group Alea supported by CNRS: the yearly meetings are held at Luminy near Marseilles, and the participation nears 80 every year.

For its 35th Special Anniversary Issue, the journal Discrete Mathematics reprinted a selection of 23 papers published in its 37 years, (Volume 306, Issue 10–11). Among them is Philippe Flajolet's 1980 "Combinatorial Aspects of Continued Fractions".

Philippe Flajolet is also an external member of the Recruiting Committee for computer science at the École polytechnique. Since 2005, Philippe Flajolet has assumed the somewhat heavy responsibility of chairing the Scientific Committee for Mathematics of the newly formed National Research Agency (ANR), which implied the heavy responsibility of launching a programme of some 5 million Euros. He is a member of the French Academy of Sciences and of the Academia Europaea.

Mireille Régnier has been on the program committee of the satellite meeting on Regulation of RECOMB'06. She participated to the organization of the new computer science option at the French *agrégation* of mathematics. She served in the committee in 2006.

Bruno Salvy has been on the program committee of the conference "Computational Geometry and Applications", Nice, 2006 and he is on the program committee for the conference Formal Power Series and Algebraic Combinatorics, Talca, Chile, 2008. He is organizing the working group Computer Algebra of the CNRS GDR IM ("Mathematical Informatics"). He is member of the editorial board of the Journal of Symbolic Computation and of the Journal of Algebra (section Computational Algebra).

He is a member of the recruiting committees of the University of Lille (in computer science), of the University of La Rochelle (in mathematics), and has been till August this year a member of the scientific council of the University of Versailles. At Inria Rocquencourt, he has organized the committees for post-docs and for the recruitment of researchers from other institutes and universities (détachements and délégations) together with Stéphane Gaubert.

This year, he has been a reviewer for the PhD L. Fousse (Loria, Nancy) and an external examiner of the PhD thesis of C. Beaumont (U. Canterbury, UK). He has also been in the Habilitation committee of F. Boulier (Lille).

8.2. Teaching

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

Alin Bostan, *Frédéric Chyzak*, and *Bruno Salvy* teach a course in computer algebra together with Marc Giusti (from École polytechnique). This course is part of the *Master Parisien de Recherche en Informatique* (MPRI).

Philippe Flajolet gives on average 24h/year at the Parisian Master of Research in Computer Science, on the analysis of algorithms.

Mireille Régnier gives 35h/year of courses in the Bioinformatics Master of Évry and Orsay and in École Centrale, Paris, on combinatorics and algorithms in genomics. She participated to a course on Regulation at Almaty, Kazakhstan, and was also invited to give one week of postgraduate courses in Almaty.

8.3. Participation in conferences, seminars, invitations

Alin Bostan, *Frédéric Chyzak*, and *Bruno Salvy*'s joint work with *Thomas Cluzeau* [22] has been presented at the conference ISSAC'06 (Genova, Italy).

Philippe Flajolet has been an invited speaker at the International Symposium on Theoretical Aspects of Computer Science (STACS 2006) Marseille, France; at the Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities, Nancy, France; at the conference Gascom'06 on random generation, Dijon, France; at the ACM-SIAM Symposium on Discrete algorithms (SODA'07), New Orleans.

Éric Fusy has been invited for one week in Zurich (jan. 2006), where he colaborated with the algorithms group and gave a talk on random generation of unlabeled structures at the weekly seminar. He was also invited in Berlin (june 2006) by the discrete mathematics team, where he gave two talks, at the CGC seminar on the enumeration of unlabeled maps and at the weekly seminar of discrete mathematics on the problem of counting bipolar orientations. He has presented the ANR project Flux at the Paristic meeting in Loria, Nancy.

Frédéric Meunier has given a talk at École des Ponts et Chaussées, about a paintshop problem in combinatorial optimization.

Mireille Régnier visited NIH (Moscow) for one week.

Bruno Salvy has been invited to give a talk at the Mathematical Physics seminar of the University Paris VII (on "Analytic combinatorics of connected graphs"). He presented a survey of recent work on the resolution linear differential and difference equations in Waterloo (Canada) on the occasion of S. Abramov 60th birthday and in Toulouse for the annual meeting of the Gecko project of the ANR. He has been invited to give a talk on the complexity of Gröbner bases at IMA (Minneapolis) for a workshop on "Algorithms in Algebraic Geometry". He has also given a mini-course on "Automatic Proofs of Special Functions or Combinatorial Identities" for professors in *classes préparatoires* at the *Journées X-UPS*.

8.4. Foreign Visitors

A large number of our visitors have given talks at the seminar of the project. This year, we received:

Stefan Gerhold, RISC, University of Linz, Austria; Jürgen Gerhard, MapleSoft, Waterloo, Canada; Éric Schost, London, Ontario, Canada; Antonio Cafure, Universidad de Buenos Aires, Universidad Nacional de General Sarmiento Buenos; Robin Pemantle, University of Pennsylvania, Philadelphia, PA, USA.

9. Bibliography

Year Publications

Books and Monographs

[1] P. FLAJOLET, R. SEDGEWICK. *Analytic Combinatorics*, Chapters I–IX of a book to be published by Cambridge University Press, 717p.+x, available electronically from P. Flajolet's home page., Cambridge University Press, April 2006.

Doctoral dissertations and Habilitation theses

[2] J. FAYOLLE. *Compression de données sans perte et combinatoire analytique*, Ph. D. Thesis, École doctorale en informatique, télécommunication et électronique de l'Université Pierre et Marie Curie, 2006, http://algo.inria.fr/papers/pdf/Fayolle06.pdf.

[3] F. GIROIRE. *Réseaux, algorithmique et analyse combinatoire de grands ensembles*, Ph. D. Thesis, Université Paris VI, November 2006, http://algo.inria.fr/papers/pdf/Giroire06b.pdf.

Articles in refereed journals and book chapters

[4] V. BOEVA, V. MAKEEV, D. PAPATSENKO, M. RÉGNIER. *Short fuzzy tandem repeats in genomic sequences, identification, and possible role in regulation of gene expression*, in "Bioinformatics", vol. 22, n° 6, 2006, p. 676–684.

[5] A. BOSTAN, P. FLAJOLET, B. SALVY, É. SCHOST. *Fast Computation of Special Resultants*, in "Journal of Symbolic Computation", vol. 41, n° 1, January 2006, p. 1–29.

[6] A. BOSTAN, P. GAUDRY, É. SCHOST. *Linear recurrences with polynomial coefficients and application to integer factorization and Cartier-Manin operator*, in "SIAM Journal on Computing", To appear, 2007.

[7] F. CHYZAK, M. DRMOTA, T. KLAUSNER, G. KOK. *The distribution of patterns in random trees*, in "Combinatorics, Probability & Computing", To appear, 2007.

[8] E. V. F. CONRAD, P. FLAJOLET. *The Fermat cubic, elliptic functions, continued fractions, and a combinatorial excursion*, in "Séminaire Lotharingien de Combinatoire", vol. 54, n^o B54g, 2006, p. 1–44.

[9] P. FLAJOLET, P. DUMAS, V. PUYHAUBERT. *Some exactly solvable models of urn process theory*, in "Discrete Mathematics & Theoretical Computer Science (Proceedings)", vol. AG, 2006, p. 59–118.

[10] P. FLAJOLET, É. FUSY, X. GOURDON, D. PANARIO, N. POUYANNE. *A Hybrid of Darboux's Method and Singularity Analysis in Combinatorial Asymptotics*, in "The Electronic Journal of Combinatorics", 35 pages, vol. 13, n° 1, June 2006, R103, http://www.combinatorics.org/Volume_13/PDF/v13i1r103.pdf.

[11] P. FLAJOLET, M. NEBEL, H. PRODINGER. *The scientific works of Rainer Kemp (1949–2004)*, in "Theoretical Computer Science", vol. 355, n° 3, April 2006, p. 371–381.

[12] P. FLAJOLET, W. SZPANKOWSKI, B. VALLÉE. *Hidden Word Statistics*, in "Journal of the ACM", vol. 53, n° 1, January 2006, p. 147–183.

[13] É. FUSY. *Counting d-polytopes with (d+3) vertices*, in "Electronic Journal of Combinatorics", vol. 13, n° 1, 2006, http://www.combinatorics.org/Volume_13/PDF/v13i1r23.pdf.

[14] É. FUSY, D. POULLAHON, G. SCHAEFFER. *Dissections, orientations and trees, with applications to optimal mesh encoding and to random sampling*, in "Transactions on algorithms", To appear, 2007.

[15] M. GIUSTI, G. LECERF, B. SALVY, J.-C. YAKOUBSOHN. *On Location and Approximation of Clusters of Zeroes: Case of Embedding Dimension One*, in "Foundations of Computational Mathematics", vol. 6, n° 3, July 2006, p. 1–57.

[16] C. GOMEZ, B. SALVY. *Calcul Formel*, in "Les Techniques de l'Ingénieur", 34 pages, to appear, vol. Dossier AF1460, 2007.

[17] M. RÉGNIER, M. VANDENBOGAERT. *Comparison of Statistical Significance Criteria*, in "Journal of Bioinformatics and Computational Biology", vol. 4, n° 2, 2006, p. 537–551, http://algo.inria.fr/regnier/publis/ReVa05.ps.

[18] F. TAHI, S. ENGELEN, M. RÉGNIER. *P-dcfold or How to Predict all Kinds of Pseudoknots in RNA Secondary Structures*, in "International Journal on Artificial Intelligence Tools", vol. 5, n° 14, 2005, p. 703–716.

Publications in Conferences and Workshops

[19] O. AMINI, J.-C. BERMOND, F. GIROIRE, F. HUC, S. PÉRENNES. *Design of Minimal Fault Tolerant Networks: Asymptotic Bounds*, in "Huitièmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel'06), Trégastel, France", May 2006, p. 37–40.

[20] O. BODINI, É. FUSY, C. PIVOTEAU. *Random sampling of plane partitions*, in "Gascom 2006, Dijon, France", R. PINZANI, V. VAJNOVSZKI (editors). , LE2I, 2006, p. 124–135.

[21] M. BODIRSKY, É. FUSY, M. KANG, S. VIGERSKE. *An unbiased pointing operator for unlabeled structures, with applications to counting and sampling*, in "18th ACM-SIAM Symposium on Discrete Algorithms, New Orleans", To appear, 2007.

[22] A. BOSTAN, F. CHYZAK, T. CLUZEAU, B. SALVY. *Low Complexity Algorithms for Linear Recurrences*, in "ISSAC'06", J.-G. DUMAS (editor). , ACM Press, 2006, p. 31–38.

[23] A. BOSTAN, F. CHYZAK, F. OLLIVIER, B. SALVY, É. SCHOST, A. SEDOGLAVIC. *Fast computation of power series solutions of systems of differential equations*, in "18th ACM-SIAM Symposium on Discrete Algorithms", New Orleans, January 2007, to appear, 2007.

[24] P. FLAJOLET. *The Ubiquitous Digital Tree*, in "STACS 2006", B. DURAND, W. THOMAS (editors). , Lecture Notes in Computer Science, Proceedings of 23rd Annual Symposium on Theoretical Aspects of Computer Science, Marseille, February 2006., vol. 3884, 2006, p. 1–22.

[25] É. FUSY. *Transversal structures on triangulations, with application to straight-line drawing*, in "Graph Drawing 2005", P. HEALY, N. S. NIKOLOV (editors). , Lecture Notes in Computer Science, Limerick, Ireland, September 12–14, 2005, vol. 3843, Springer, 2006, p. 177–188.

[26] É. FUSY. *Straight-line drawing of quadrangulations*, in "Graph Drawing 2006", Karlsruhe, Germany, September 18–20, 2006, to appear, Springer-Verlag, 2007.

[27] É. FUSY, F. GIROIRE. *Estimating the number of active flows in a data stream over a sliding window*, in "ANALCO'07 (New Orleans)", To appear, 2007.

[28] F. GIROIRE. *Directions to use Probabilistic Algorithms for Cardinality for DNA Analysis*, in "Journées Ouvertes Biologie Informatique Mathématiques (JOBIM 2006)", July 2006, p. 3–5.

Internal Reports

[29] A. BOSTAN, F. MORAIN, B. SALVY, É. SCHOST. *Fast algorithms for computing isogenies between elliptic curves*, Submitted, Technical report, n° 91441, Inria, 2006, https://hal.inria.fr/inria-00091441.

Miscellaneous

[30] B. BESSON. *Matrices poids-position : principes et algorithmes*, Mémoire de Master 2, Université Paris XI, 2006.

[31] M. BODIRSKY, É. FUSY, M. KANG, S. VIGERSKE. *Enumeration of unlabeled outerplanar graphs*, February 2006, Submitted.