

*Projet ALGO**Algorithmes**Rocquencourt*

THÈME 2B



*R*apport
d'Activité

2000

Table des matières

1	Composition de l'équipe	2
2	Présentation et objectifs généraux	3
3	Fondements scientifiques	4
3.1	Analyse d'algorithmes	4
3.2	Calcul formel	6
3.3	Algorithmique des séquences	8
3.4	Algorithmique et modélisation des réseaux	9
4	Domaines d'applications	10
4.1	Panorama	10
5	Logiciels	11
6	Résultats nouveaux	12
6.1	Analyse d'algorithmes	12
6.2	Calcul formel	16
6.3	Algorithmiques des séquences	17
6.4	Algorithmique et modélisation des réseaux	18
7	Contrats industriels (nationaux, européens et internationaux)	20
7.1	Algorithmique et modélisation des réseaux	20
7.2	Calcul formel	20
8	Actions régionales, nationales et internationales	20
8.1	Actions nationales	20
8.2	Actions financées par la Commission Européenne	21
8.3	Relations bilatérales internationales	21
8.4	Accueils de chercheurs étrangers	21
9	Diffusion de résultats	22
9.1	Animation de la communauté scientifique	22
9.2	Enseignement universitaire	23
9.3	Participation à des colloques, séminaires, invitations	23
10	Bibliographie	25

1 Composition de l'équipe

Responsable scientifique

Bruno Salvy [DR]

Responsable permanent

Philippe Flajolet [DR]

Assistante de projet

Virginie Collette [TR]

Personnel Inria

Frédéric Chyzak [CR]

Mireille Régnier [DR]

Philippe Robert [DR]

Vincent Dumas [MC, Université de Bordeaux, en délégation depuis le 1er septembre 2000]

Collaborateurs extérieurs

Jean-François Dantzer [Prag à l'Université de Versailles St-Quentin]

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

Pierre Nicodème [CR CNRS, Université d'Évry]

Mathieu Raffinot [CR CNRS, Université d'Évry]

Michèle Soria [professeur, Université Paris VI]

Brigitte Vallée [professeur, Université de Caen]

Doctorants

Cyril Banderier [Université de Paris VI]

Julien Clément [Université de Caen (temps partiel)]

Mostapha Haddani [Université de Versailles St-Quentin]

Marianne Durand [ENS Paris]

Mathias Vandenbergert [Université de Bordeaux]

Elchanan Mossel [Université Hébraïque, Israël, jusqu'à mi-mars 2000]

Stagiaires

Bruno Boussard [Université d'Orsay, juillet à août 2000]

Marni Mishna [Université du Québec à Montreal]

2 Présentation et objectifs généraux

L'objectif global du projet ALGO est l'analyse et l'optimisation fines de systèmes complexes discrets présentant une forte composante aléatoire. De nombreux problèmes de grande taille rentrent dans ce cadre tels l'étude quantitative d'algorithmes probabilistes sur des structures discrètes, ou l'optimisation de l'allocation de ressources dans les réseaux de communication. La réalisation de cet objectif passe par la compréhension en profondeur de l'aléa discret et des problèmes de nature mathématique que pose sa quantification. Cela suppose de dégager des méthodes générales pour obtenir des résultats présentés de manière exacte ou asymptotique. Ces résultats fournissent alors des informations très précises sur le comportement qualitatif ou quantitatif des systèmes étudiés.

Étant donné le caractère très systématique de l'approche poursuivie, des méthodes de décision réalisables en calcul formel font aussi partie des objectifs du projet. Cette approche est un moteur puissant de renouvellement qui conduit à la révision d'approches classiques dans le domaine des fonctions spéciales et des développements en séries. L'objectif est de disposer d'une algorithmique fiable et complète pour de grandes classes de problèmes précisément caractérisés ; voir déjà les bibliothèques GFUN et MGFUN assez largement utilisées dans la communauté combinatoire et présentes dans la dernière version de Maple. Les résultats sont susceptibles de nombreuses applications bien au delà du domaine de la modélisation combinatoire : ainsi une meilleure intégration des fonctions spéciales au calcul formel est visée, ce qui s'applique à de larges classes de problèmes dans les sciences de l'ingénieur.

Thématique de recherche

Le projet ALGO se donne comme objectif l'analyse en profondeur de l'aléa combinatoire et la recherche de ses *lois* générales. Ce thème est voisin par ses objectifs, mais dual par ses méthodes, de la modélisation des systèmes informatiques, laquelle repose de manière prédominante sur des mathématiques *a priori* continues. Ici, nous sommes dans le domaine des mathématiques discrètes. La combinatoire étant par définition l'étude des objets finis discrets, nous visons à développer une approche globale que l'on pourrait qualifier de "combinatoire statistique" par analogie avec la physique statistique. Le but est de décrire le comportement macroscopique visible de l'objet étudié, comme par exemple, le temps d'exécution sur un grand "ensemble" d'un certain algorithme, ou encore l'évolution d'un grand réseau vers ses régimes fondamentaux.

Il s'agit ainsi de caractériser les propriétés attendues (en moyenne, en probabilité ou en distribution) d'objets obéissant à des règles de combinaison finies, mais constituant de très grands ensembles. Ces situations se rencontrent sans cesse en informatique, un tri de seulement 100 éléments met en jeu 10^{158} configurations possibles, qui obéissent, avec une écrasante probabilité, à des règles fort précises. Les problèmes d'aléa combinatoire interviennent de manière essentielle en algorithmique. La conception de la plupart des algorithmes efficaces se fonde

naturellement sur les cas attendus, en moyenne ou en probabilité, plutôt que sur une analyse pessimiste qui doit être réservée à des contraintes de type “temps réel”.

Le rôle de l'analyse de l'aléa combinatoire est renforcé par l'importance croissante des algorithmes dits “randomisés” (bien formalisés par Karp et Rabin depuis une vingtaine d'années) où il s'avère payant d'introduire volontairement le hasard dans le calcul. Ainsi, les tables de hachage constituent un substitut souvent très efficace aux arbres de recherche, les signatures accélèrent considérablement la recherche textuelle, les “skip lists” remplacent graduellement les arbres équilibrés dans de nombreuses applications. Parmi d'autres applications célèbres de cette “aléatorisation”, on peut citer la construction de cryptosystèmes à clefs publiques qui utilise de manière très sûre des tests probabilistes ; ou encore la conception de protocoles de communication, faisant suite à Ethernet, mais avec une utilisation mieux contrôlée de l'aléa qui permettent de stabiliser et d'acheminer des trafics plus importants dans les réseaux. La problématique de l'algorithmique randomisée est poursuivie par exemple avec succès par le projet PRISME de Sophia-Antipolis dans le domaine de l'algorithmique géométrique et par le projet HIPERCOM de Rocquencourt dans le domaine des télécommunications.

Le développement récent du *calcul formel* a éclairé tout l'intérêt de cette approche généraliste. Il est apparu, en effet, vers le tournant des années 1990 qu'il était possible de décider mathématiquement de nombreuses propriétés de l'aléa combinatoire, ce par un calcul de nature essentiellement formelle. Au sein du projet, cela a donné lieu aux thèses de B. Salvy et Paul Zimmermann (en 1991). Le programme de recherche correspondant à ces aspects est loin d'être achevé. De nombreux problèmes de portée générale en calcul formel ont été mis en évidence. Citons principalement ici, comme directions nouvelles, l'asymptotique automatique (B. Salvy), les méthodes mixtes symboliques numériques (B. Salvy), et la preuve automatique d'identités combinatoires (F. Chyzak, B. Salvy).

C'est dans ce contexte qu'ont pu être résolues au fil des ans diverses conjectures correspondant à des analyses précises d'algorithmes tels que le dimensionnement en hachage dynamique (M. Régnier, P. Flajolet), la performance des arbres quadrants pour la recherche multidimensionnelle (P. Flajolet, B. Salvy), ou le comportement probabiliste des meilleures méthodes de recherche de motifs (M. Régnier). D'autres applications typiques sont constituées par les algorithmes d'estimation probabiliste en bases de données, l'allocation mémoire partagée (P. Flajolet), les protocoles en arbre (P. Flajolet et Philippe Jacquet), et le comportement probabiliste de grands réseaux (P. Robert). Les travaux du groupe sont ainsi assez largement repris dans la littérature scientifique spécialisée (livres ou articles). Ils nous valent de nombreuses coopérations internationales avec des universités comme celles de Barcelone, Newcastle, Princeton, Purdue, Stanford, Vienne, Washington, Waterloo, etc.

3 Fondements scientifiques

3.1 Analyse d'algorithmes

Participants : Cyril Banderier, Julien Clément, Marianne Durand, Philippe Flajolet, Mireille Régnier, Bruno Salvy, Michèle Soria, Brigitte Vallée.

Mots clés : analyse d'algorithmes, analyse combinatoire, combinatoire analytique,

structures aléatoires discrètes, arbre d'index, algorithme de hachage, algorithme géométrique, loi limite.

Résumé :

L'analyse d'algorithmes vise à une quantification précise des principaux algorithmes et structures de données de l'informatique. Elle repose sur une étude approfondie de l'aléa discret. Une approche originale développée au sein du projet se fonde sur la "combinatoire analytique". Les applications en sont les principaux problèmes de recherche de l'information dans de grands volumes de données et l'algorithmique de la communication.

L'analyse d'algorithmes ou de structures de données dépend très étroitement d'une évaluation et d'une quantification précise de l'aléa discret. Il s'agit en effet de caractériser le comportement des principaux paramètres de structures combinatoires d'arbres, de mots, de graphes ou d'allocations aléatoires. Dans ce domaine, les travaux du projet ALGO s'appuient sur une approche originale fondée sur :

- les méthodes symboliques en analyse combinatoire qui permettent de disposer de méthodes systématiques de mise en équation par séries génératrices ;
- les méthodes asymptotiques complexes fondées notamment sur l'analyse de singularité.

L'application de ces idées aux séquences textuelles ou génétiques fait l'objet de la section 3.3. L'automatisation de ces méthodes visant à l'analyse complète de modèles complexes est traitée à la section 3.2.

Traditionnellement, l'analyse d'algorithmes propose des prédictions quantitatives de comportement moyen, c'est-à-dire des analyses d'espérances de coûts. Une évolution importante qui commence au début des années 1990 avec la thèse de doctorat ès-sciences de Michèle Soria consiste à aborder des études en distribution. Ceci permet non seulement de connaître les comportements moyens attendus mais encore de prédire les profils de coûts sous la forme de distributions limites ou de bornes de grandes déviations. En d'autres termes, on parvient depuis peu à quantifier très précisément les compromis de type risque-efficacité. Du point de vue mathématique, les travaux établissent un pont intéressant entre approches probabilistes classiques et approches issues de la combinatoire analytique. Une monographie intitulée "Analytic Combinatorics" par P. Flajolet et Robert Sedgewick est en préparation (500 pages sont déjà rédigées et publiées en rapport de recherche INRIA). Cette monographie fait partie d'un programme de recherche visant à établir un corpus cohérent de méthodes pour le domaine de la combinatoire analytique et pour la quantification de l'aléa discret. Les méthodes qui sous-tendent ce programme ont de nombreuses applications en analyse d'algorithmes, parmi lesquelles sont traités spécifiquement dans le projet : les structures de données fondamentales comme les tables de hachage, les arbres de recherche uni- ou multi-dimensionnels (voir la section 6.1) ; l'algorithmique de la recherche textuelle et du traitement de séquences génomiques (voir la section 6.3) ; les arbres digitaux liés à des problèmes de nature très générale en théorie analytique de l'information (voir les sections 6.1). Ces méthodes inspirent également une partie des recherches en calcul formel (section 6.2) par les problèmes que pose l'automatisation de la manipulation de grands modèles combinatoires.

3.2 Calcul formel

Participants : Frédéric Chyzak, Marni Mishna, Bruno Salvy.

Mots clés : échelles asymptotiques, génération aléatoire, fonctions spéciales, élimination polynomiale, bases de Gröbner.

Résumé : *Le projet ALGO développe des algorithmes de calcul formel qui permettent : le traitement de modèles combinatoires récurrents ; l'analyse asymptotique automatique de nombreuses classes de problèmes ; le traitement de fonctions spéciales et de suites combinatoires de manière unifiée, et notamment leur intégration ou leur sommation. Ces algorithmes sont dans la plupart des cas validés par des implantations. Leur développement est motivé par l'automatisation d'une approche combinatoire à l'analyse d'algorithmes.*

Les trois étapes fondamentales de l'analyse d'algorithmes telle qu'elle est pratiquée au projet ALGO sont la modélisation combinatoire, la manipulation de séries génératrices et l'analyse asymptotique. Chacune de ces étapes requiert des capacités de calcul symbolique importantes, tant pour l'application des méthodes symboliques que pour l'expérimentation. Ce besoin explique l'importance de l'activité en calcul formel au sein du projet. L'objectif à long terme est de systématiser et d'automatiser ces trois étapes.

Au cours des années ont été élaborés de nombreux algorithmes et programmes. Les domaines mathématiques permettant l'expression et la preuve de nos algorithmes sont la combinatoire, l'analyse asymptotique, l'algèbre différentielle (corps de Hardy et anneaux de polynômes non commutatifs). Notre travail en calcul formel est ainsi complémentaire de celui des projets SAGA et CAFE consacrés aux systèmes polynomiaux et aux équations différentielles.

Comme pour l'analyse d'algorithmes, l'approche du projet aux algorithmes du calcul formel est globalisante et unificatrice. La résolution de problèmes appliqués mettant en œuvre de la combinatoire ou de l'analyse d'algorithmes est abordée à un niveau de généralité qui permet le développement d'une algorithmique de portée large. Ainsi, les travaux sur la combinatoire fournissent des générateurs aléatoires efficaces susceptibles de nombreuses applications ; les travaux sur l'analyse d'algorithmes ont abouti au développement d'une algorithmique d'échelles asymptotiques très générale, dont le besoin s'était fait sentir en intégration numérique et en physique mathématique. Les outils développés dans le projet touchent maintenant un public assez large d'utilisateurs du calcul formel intéressés tant par la combinatoire que par les manipulations de séries ou par l'analyse asymptotique. La diversité de ce public est encore accrue par la présence de certains de nos programmes dans les bibliothèques du système de calcul formel MAPLE, l'un des deux principaux systèmes généralistes actuellement disponibles.

On peut distinguer trois grandes directions de travail dans notre activité en calcul formel : les structures combinatoires, les suites et fonctions spéciales et l'asymptotique automatique.

Structures combinatoires Un langage de description généralisant les grammaires *context-free* permet d'exprimer des objets aussi divers que permutations, arbres binaires, arbres généraux, partitions d'entiers ou d'ensembles, graphes fonctionnels ou molécules chimiques, par

exemple carbures ou alcools. À partir d'une description de structure décomposable, il est possible de (i) compter efficacement le nombre d'objets d'une certaine taille répondant à la spécification; (ii) produire des fonctions de génération aléatoire uniforme de faible complexité — utiles pour des tests statistiques; (iii) produire des fonctions de génération exhaustive de ces objets — utiles pour des tests de robustesse de procédures; (iv) produire des itérateurs, c'est-à-dire des fonctions permettant d'accéder successivement à tous les objets d'une certaine taille, mais sans les stocker tous en mémoire; (v) calculer des équations satisfaites par les séries génératrices d'énumération de ces objets — utiles pour la phase d'analyse asymptotique.

Le programme $\Lambda\Gamma\Omega$ réalisé au début des années 90 par B. Salvy et P. Zimmermann fournissait une partie de ces fonctionnalités, mais sa portabilité et ses fonctionnalités étaient limitées par l'usage de CAML en conjonction avec MAPLE. L'objectif poursuivi depuis plusieurs années est de tirer parti de l'expérience acquise avec $\Lambda\Gamma\Omega$ pour réaliser une version (COMBSTRUCT) entièrement intégrée en MAPLE, en mettant l'accent sur la modularité, la robustesse et la souplesse d'emploi.

Suites et fonctions spéciales Selon l'origine combinatoire du problème, les séries génératrices que l'on est amené à étudier peuvent être données sous des formes diverses. Elles peuvent être connues sous forme explicite mais elles sont le plus souvent décrites sous forme implicite par une ou plusieurs équations, fonctionnelles, différentielles ou aux différences. De même, leurs coefficients peuvent vérifier des récurrences de natures diverses. Manipuler ces fonctions définies implicitement nécessite des innovations théoriques, ainsi qu'un important effort d'implantation. Ce thème de recherche touche aux fondements du calcul symbolique, où il apparaît qu'il est paradoxalement souvent plus facile de traiter une fonction lorsqu'elle est représentée comme solution d'équations que lorsqu'elle est représentée sous forme close. En particulier, les questions de simplification et de formes normales qui sont une des difficultés majeures rencontrées par l'utilisateur trouvent une bien meilleure réponse dans ce contexte.

Le cas des solutions d'équations différentielles ou de récurrences linéaires attire beaucoup l'attention de la communauté de combinatoire et de calcul formel. De nombreuses suites et fonctions spéciales sont définies par de telles équations, qui bénéficient d'une algorithmique très riche, implantée en grande partie dans le *package* GFUN développé par B. Salvy et P. Zimmermann (projet POLKA). Le pendant multivarié de ces travaux est extrêmement fructueux, puisqu'il permet d'envisager une approche algorithmique de nombreux calculs avec des polynômes orthogonaux, des fonctions spéciales, ou plus généralement des fonctions, suites, séries ou distributions définies par un *système* d'équations linéaires à coefficients polynomiaux. Le travail mené par F. Chyzak considère des systèmes d'opérateurs linéaires qui peuvent être différentiels, aux différences, aux q -différences, ou de nombreux autres types. Il s'avère que les opérations d'addition, de produit, d'intégration ou de sommation peuvent s'effectuer algorithmiquement par le biais de techniques d'élimination sophistiquées dans des algèbres non commutatives d'opérateurs linéaires. Le *package* MGFUN développé par F. Chyzak valide et motive les recherches théoriques correspondantes.

Séries et échelles asymptotiques Les besoins de la combinatoire analytique en matière de développements asymptotiques dépassent les capacités actuelles des systèmes de calcul formel.

En effet, les calculs de coûts moyens et plus encore de variance donnent systématiquement lieu à des annulations non seulement dans les premiers termes des développements mais aussi dans l'ordre de grandeur exponentiel des croissances. La construction automatique des échelles asymptotiques nécessaires et le calcul avec ces échelles pose de nombreux problèmes sur lesquels le calcul formel est en progrès rapide. Les premiers travaux sur ce sujet datent des années 90. En 1988, Gaston Gonnet et Keith Geddes (créateurs du système MAPLE) proposent un modèle permettant de traiter des formules de complexité proche de la formule de Stirling. Puis en 1990, John Shackell (University of Canterbury) publie un algorithme qui permet de déterminer *de manière garantie* la limite des fonctions exp-log (fonctions de base de l'asymptotique). L'année suivante, la thèse de B. Salvy propose une première implantation de développements asymptotiques dans des échelles asymptotiques générales. Depuis lors, grâce à une série de travaux en commun de B. Salvy et J. Shackell, des classes de plus en plus vastes de fonctions ont reçu un algorithme permettant le calcul asymptotique de manière garantie. Ces progrès n'ont pas toujours été suivis d'implantation, mais une nouvelle structure de données pour les développements asymptotiques, les multiséries, a vu le jour récemment. L'algorithmisation de cette structure de données conduit à revisiter et prolonger les algorithmes anciens, et mène à la réalisation de prototypes permettant déjà de résoudre des problèmes du niveau de la recherche. Cette structure de données a fait l'objet d'une proposition aux développeurs du système MAPLE. Là encore, on s'attaque à une brique essentielle des systèmes de calcul formel, puisque les séries sont la structure de données efficace utilisée pour manipuler les polynômes, eux-mêmes à la base de nombreuses structures de données.

3.3 Algorithmique des séquences

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

Mots clés : combinatoire des mots, séquences, recherche de motifs, génome.

Résumé : *L'objet des recherches sur les séquences est la conception de nouveaux algorithmes, l'obtention de la complexité moyenne de ces algorithmes et l'application à l'algorithmique de certains résultats statistiques. Plus généralement, nous développons une théorie analytique de l'information qui s'appuie sur la combinatoire, les probabilités et l'analyse.*

L'algorithmique des séquences ou objets textuels couvre des domaines d'application variés (compression des données textuelles et séquences biologiques). Ce sujet comprend d'abord des recherches algorithmiques. Il s'agit de trouver efficacement un motif ou un ensemble de motifs dans un texte. Cet ensemble peut en particulier être l'ensemble des mots voisins d'un mot donné, à un nombre d'erreurs près, où le type et le nombre d'erreurs autorisées sont déterminés par l'application. Une nouvelle classe d'algorithmes dits d'extraction de motifs, émerge dans les recherches sur le génome. On extrait des motifs à signification biologique particulière, non connus à l'avance. Ce sont par exemple des mots à fréquence exceptionnelle : surreprésentés ou sousreprésentés pour un modèle probabiliste donné. On peut aussi rechercher une expression

régulière caractéristique d'une famille donnée. Ces algorithmes testent sur une même séquence un grand nombre de candidats potentiels et la rapidité du calcul statistique, et sa précision, apparaissent essentielles.

D'un point de vue méthodologique, des théorèmes probabilistes trouvent des applications naturelles dans l'étude des séquences. Plus précisément, nous avons mis en évidence différents types de processus de renouvellement, la loi limite étant généralement gaussienne; le calcul effectif des paramètres de coût peut être très délicat et les outils combinatoires et analytiques permettent pour cette classe de problèmes les calculs effectifs des distributions. Les formules obtenues sont aussi valables dans le domaine fini. Enfin, l'étude des grandes déviations, en cours, est très prometteuse. Parallèlement, nous développons des outils de calcul dans le modèle probabiliste markovien. La complexité des évaluations de performances dans le cas markovien provient du nombre de cas différents à considérer. Nous définissons pour chaque problème des langages caractéristiques dont la contribution au coût total de l'algorithme est calculable. Ceci équivaut à une agrégation des états de l'automate associé. Les résultats s'appliquent à la compression et à la recherche de motifs avec erreurs ou de motifs exceptionnels dans les textes (DosDNA, reconnaissance de gènes ou de signaux de régulation, etc.). Ils permettent aussi d'établir les domaines d'efficacité des différents algorithmes de recherche de similarités dans des bases de données protéiques.

3.4 Algorithmique et modélisation des réseaux

Participants : Jean-François Dantzer, Mostafa Haddani, Philippe Robert.

Mots clés : grands réseaux, méthode de renormalisation, vitesse de convergence.

Résumé : *L'étude quantitative des modèles probabilistes de réseaux, complexe dès lors que le nombre de nœuds est grand, est envisagée du point de vue des techniques de renormalisation de la physique statistique. L'état du réseau est renormalisé avec un paramètre (la taille, la capacité du réseau, l'intensité du trafic) qui tend vers une valeur critique. Asymptotiquement, l'étude du comportement du réseau se ramène à la résolution d'équations différentielles quasi-déterministes.*

Le cadre général de cette recherche concerne les propriétés de renormalisation des réseaux de communication. Le processus de Markov décrivant l'état d'un réseau est en général complexe, même si la distribution à l'équilibre de celui-ci est connue. La combinatoire des expressions ne permet pas une évaluation qualitative de ces réseaux, pour les problèmes de dimensionnement notamment, lorsque leur nombre de nœuds est significatif. Une méthode intéressante, issue de la physique des particules, consiste à renormaliser le processus à la fois en temps et en espace par un petit paramètre ε et faire tendre celui-ci vers 0. Un processus limite ainsi obtenu conserve les caractéristiques essentielles du réseau, schématiquement la partie bruit autour des trajectoires est éliminée. Le processus limite obtenu de cette façon est quasi-déterministe, les points de discontinuité de la dynamique conservant une part d'aléatoire. Toute la difficulté de l'étude consiste à identifier et caractériser les limites possibles. Il peut y avoir plusieurs limites et les équations "limites" peuvent présenter des solutions pathologiques qu'il convient

d'éliminer (les physiciens le font avec des arguments d'entropie). La renormalisation présente l'intérêt de mettre en évidence les modes de fonctionnement fondamentaux. Ce programme d'étude a déjà largement été entamé dans la thèse de V. Dumas et développé par J.-F. Dantzer dans le cadre des réseaux multi-classe et par M. Haddani dans une thèse en cours sur les questions d'allocation de bande passante.

Les problèmes de vitesse de convergence constituent un aspect important de notre étude. Ils se situent en amont des questions de renormalisation. Schématiquement, l'étude d'un système donné peut se décrire de la façon suivante : étude de la renormalisation, propriétés gaussiennes autour des trajectoires renormalisées, existence d'un principe de grandes déviations, et enfin étude de la convergence à l'équilibre des modèles considérés. Jusqu'à présent, les résultats obtenus dans ce domaine sont principalement des estimations du taux de convergence exponentielle par rapport au temps, pour obtenir des bornes du type $Ke^{-\lambda_2 t}$. La constante K , qui est généralement négligée, joue cependant assez souvent un rôle majeur dans la convergence à l'équilibre. Par exemple, l'étude de la seconde valeur propre λ_2 suggère souvent un temps de relaxation (le temps pour que la distance à l'équilibre soit inférieur à $1/e$) de l'ordre de $1/\lambda_2$. Il est cependant très fréquent que l'ordre de grandeur du véritable temps de relaxation soit complètement différent. L'accumulation de valeurs propres au voisinage de λ_2 semble être à l'origine de ces phénomènes encore mal compris. Notre objectif principal dans ce cadre est de donner des *bornes* aussi simples et précises que possible sur la distance entre l'état d'un système à un instant donné et ce même système à l'équilibre. L'intérêt de tels résultats est de pouvoir donner une description du comportement transitoire d'un système qui n'est pas à l'équilibre. Si l'état d'équilibre a , en général, une expression difficile à utiliser, le cas transitoire est en revanche pratiquement toujours inconnu. Ce type d'étude a été mené avec succès, avec des techniques d'analyse de Fourier. Les marches aléatoires sur des graphes ayant une structure assez régulière s'étudient avec ces méthodes, voir par exemple le livre de Diaconis "*Group representations in probability and statistics*" sur ce sujet. Les techniques envisagées pour cette étude sont essentiellement probabilistes. Notre objectif est d'étudier dans un premier temps des réseaux simples et de dégager des méthodes d'investigation. Dans ce cadre aussi, nous avons recours à une variable de renormalisation qui permet de dégager le mode majeur de convergence. Outre les bornes explicites sur la distance à l'équilibre, ce type d'étude permet de définir, dans certaines situations, le temps d'atteinte de l'équilibre ; avant ce temps, il est possible de trouver un état initial pour lequel la distance est maximale, et après la distance est nulle.

4 Domaines d'applications

4.1 Panorama

Participants : Jean-François Dantzer, Mostapha Haddani, Philippe Robert.

Résumé : *Les applications de nos travaux sur les structures combinatoires sont la modélisation et l'étude des systèmes discrets complexes et des réseaux de télécommunication.*

Les applications visées par l'analyse d'algorithmes sont la conception de méthodes d'accès rapide à des informations structurées, une algorithmique rapide du calcul formel, le traitement statistique des séquences biologiques, et les protocoles de communication.

Nos domaines de recherche en calcul formel sont : les structures combinatoires, les suites et fonctions spéciales et l'analyse asymptotique. Les applications de nos travaux sur les structures combinatoires sont la modélisation et l'étude de systèmes discrets complexes. Nos résultats sur les suites et fonctions spéciales débouchent sur la manipulation par le calcul formel de fonctions spéciales intervenant de manière classique en physique mathématique et en mécanique. Nos travaux sur l'asymptotique devraient permettre à long terme de faire le pont entre le calcul numérique et le calcul formel : le calcul numérique, robuste en l'absence de singularités pourrait être complété par une étude formelle fine au voisinage des singularités débouchant sur de la production de code numérique robuste dans ces zones sensibles.

Les travaux sur les réseaux sont naturellement liés aux problèmes de dimensionnement, de routages des messages, ou encore d'optimisation des paramètres de qualité de service (délai, taux de rejet, etc.). Les grandes tailles des réseaux, les hauts débits utilisés et l'hétérogénéité des trafics (vidéo, voix, données) qui y circulent rendent de plus en plus nécessaire une utilisation accrue de modèles mathématiques.

5 Logiciels

La bibliothèque COMBSTRUCT a été conçue et développée par le projet ALGO (P. Flajolet, M. Mishna, B. Salvy, Eithne Murray) en liaison avec le projet POLKA de Nancy (P. Zimmermann). Elle fait l'objet d'une collaboration régulière avec les groupes de Waterloo (Université et Compagnie WMI) et une version assez récente est intégrée au système Maple. Elle permet actuellement la génération aléatoire ou exhaustive, le calcul automatique de dénombrements et de séries génératrices, et est à ce titre une aide de portée générale pour la simulation et le test systématique de modèles combinatoires. Disons qu'en l'état actuel, et sur son créneau, son expertise est de l'ordre de celle d'un étudiant en début de 3ème cycle. La dernière version de COMBSTRUCT comporte une importante extension à une classe de grammaires attribuées, développée par M. Mishna, qui permet de calculer automatiquement les séries génératrices de coût d'une large classe d'algorithmes. La version actuelle représente environ 8 500 lignes de code (300 ko).

Le *package* GFUN développé par B. Salvy et P. Zimmermann (projet POLKA) fournit de nombreux outils de manipulations de suites et de fonctions, à commencer par une fonction de traduction qui part de la forme close d'une fonction et produit une équation différentielle linéaire dont cette fonction est solution (lorsqu'une telle équation existe). Cette fonctionnalité, qui effectue précisément le chemin inverse de celui vers lequel se précipitent nombre d'utilisateurs, permet ensuite de calculer des développements en série de manière plus rapide qu'avec la forme close ; elle permet également la localisation des singularités et le calcul des comportements au voisinage des singularités. Le *package* GFUN comporte actuellement environ 3 300 lignes de code Maple. Il a fait l'objet d'une revue très positive dans *Computing Reviews* et est incorporé au *superseeker* de N. Sloane aux Bell Laboratories, accessible sur le Web, qui détermine de nombreuses suites d'après leurs premiers termes. Une extension de GFUN pour

produire automatiquement des procédures efficaces d'évaluation numérique en précision arbitraire fera partie de la prochaine version.

La bibliothèque GDEV pour l'asymptotique des fonctions génératrices développée par B. Salvy a été entièrement remise à jour. Ceci a permis la mise au point par Stéphanie Petit d'une mini-encyclopédie des structures combinatoires reposant sur la combinaison de COMBSTRUCT, GFUN et GDEV et dotée d'un mode d'interrogation via le Web à l'adresse <http://algo.inria.fr/encyclopedia>.

Les algorithmes développés par F. Chyzak sont implantés dans une bibliothèque Maple du nom de MGFUN partiellement intégrée dans la distribution grand public de Maple. Depuis quelques versions, les utilisateurs de Maple disposent ainsi d'une bibliothèque pour la manipulation d'opérateurs linéaires, ainsi que d'une nouvelle bibliothèque pour calculer des bases de Gröbner, capable de traiter aussi bien des polynômes que des opérateurs non commutatifs. C'est désormais cette dernière qui est employée pour réaliser l'élimination polynomiale dans tout le logiciel Maple. Par ailleurs, la simple présence dans Maple de routines de calcul sur les opérateurs linéaires a ouvert la voie à l'implantation, jusqu'alors impossible, de toute une nouvelle génération d'algorithmes pour la sommation et l'intégration symboliques, et plus généralement d'algorithmes récents pour la manipulation de représentations implicites de suites et fonctions spéciales. Suite à l'intégration de MGFUN, la société WMI qui développe Maple a accentué son effort dans cette direction. L'ensemble de la réalisation logicielle correspondant à la bibliothèque MGFUN est constitué de 14 000 lignes de code (520 Ko), accompagné d'une quantité équivalente de documentation et de jeux de tests ; elle apporte plus d'une cinquantaine de nouvelles fonctions à l'utilisateur. Afin de rendre l'utilisation de MGFUN plus transparente aux utilisateurs dont les problèmes ne requièrent pas toute la puissance du *package*, une extension de MGFUN a été réalisée en collaboration avec Cyril Germa (stagiaire dans le projet). Cette extension détermine automatiquement l'algèbre et les opérateurs avec lesquels effectuer les calculs de somme ou d'intégrale demandés par l'utilisateur. Le *package* MGFUN a par ailleurs été choisi pour ses primitives de calcul de bases de Gröbner pour le *package* DESING développé au RISC (université de Linz, Autriche) et disponible sur le Web (<http://www.risc.uni-linz.ac.at/projects/basic/adjoints/blowup/>).

L'ensemble des *packages* ci-dessus est maintenant intégré en une unique bibliothèque du nom d'*algolib*, et diffusé par les pages web du projet ALGO. Leur installation par les utilisateurs en est ainsi grandement simplifiée, tout en permettant une meilleure interaction entre les *packages*.

6 Résultats nouveaux

6.1 Analyse d'algorithmes

Participants : Cyril Banderier, Julien Clément, Marianne Durand, Philippe Flajolet, Mireille Régnier, Bruno Salvy, Michèle Soria, Brigitte Vallée.

L'analyse d'algorithmes et de structures de données dépend très étroitement d'une évaluation et d'une quantification précise de l'aléa discret. Il s'agit en effet de caractériser le comportement des principaux paramètres de structures combinatoires d'arbres, de mots, de graphes

ou d'allocations aléatoires. Dans ce domaine, les travaux du projet ALGO s'appuient sur une approche originale et systématique qui a été baptisée "Combinatoire Analytique". La combinatoire analytique repose sur deux ingrédients mathématiques essentiels, dont le développement a largement pris place dans le projet :

- les méthodes symboliques en analyse combinatoire et analyse d'algorithmes qui permettent de disposer de principes systématiques de mise en équation par séries génératrices ;
- les méthodes asymptotiques complexes fondées notamment sur l'analyse de singularité et les transformations intégrales.

Les recherches du projet ALGO se placent dans ce cadre général, et la partie décrite ici constitue le "tronc commun" à de nombreux autres travaux conduits dans l'équipe. Ces études soutiennent donc les applications aux séquences textuelles ou génétiques (section 6.3) ainsi qu'à certains des problèmes de modélisation de réseaux (voir la section 6.4). Enfin, l'automatisation de méthodes visant à l'analyse complète de modèles complexes est intimement liée au calcul formel et traitée à la section 6.2.

Combinatoire analytique. Un premier lot de travaux conçus ou publiés en l'an 2000 a trait à l'approfondissement des méthodes fondamentales de la combinatoire analytique. Deux aspects sont couverts : l'aspect combinatoire qui repose sur une élaboration des méthodes de dénombrement, et l'aspect analytique, c'est-à-dire asymptotique, qui permet d'extraire des prédictions quantitatives à la fois précises et simples dans leurs formes.

En l'an 2000, plusieurs travaux ont été consacrés à dégager une classe d'universalité (terme issu de la physique statistique) en combinatoire analytique. Il s'agit d'analyser des "transitions de phase" dont on ressent de plus en plus en plus la présence par exemple dans les phénomènes de seuil observés en optimisation combinatoire (avec régions "faciles" et "difficiles" séparées brusquement). Jusqu'ici, de nombreux phénomènes échappaient à la quantification. Il a été mis en évidence au sein du projet le rôle important de la coalescence de cols conduisant à des phénomènes décrits par fonctions d'Airy très différents des habituels comportement gaussiens ; voir les travaux [23, 36] et des recherches en cours de P. Flajolet, B. Salvy et Gilles Schaeffer (Projet POLKA, INRIA Nancy). Les domaines visés sont ceux de la coalescence dans les tables de hachage, de la connectivité dans les graphes aléatoires, ou encore de la multiconnectivité dans les cartes planaires aléatoires. Dans un domaine voisin, un article récent [25] (en liaison avec l'Université de Barcelone) montre l'applicabilité du cadre général de la combinatoire analytique au dénombrement de configurations de base de l'algorithmique géométrique telle que développée au Projet PRISME de l'INRIA Sophia-Antipolis. Il s'agit ici de l'étude des diagrammes de cordes, lesquels apparaissent tant en géométrie qu'en combinatoire ou en physique. L'article [25] montre que de nombreuses distributions de probabilités des diagrammes aléatoires succombent aux méthodes de la combinatoire analytique. Les résultats obtenus concernent par exemple le paramètre fondamental du nombre de croisement qui se relie à la riche théorie des " q -analogues", aux fractions continues, et aux analogues de polynômes orthogonaux classiques. De la sorte est établi par exemple le caractère asymptotiquement gaussien du nombre de croisements. L'intérêt de ce genre de résultat est largement méthodologique. Ces questions s'insèrent

dans l'action de recherche incitative ALCOPHYS de la direction scientifique de l'INRIA, et il est de fait que les premiers travaux dans ce domaine sont issus d'un besoin d'élucider certains développements perturbatifs liés à des changements de phase en physique statistique.

Un autre ensemble de travaux est relatif aux chemins aléatoires, objets de base de la combinatoire et des probabilités discrètes. Tout d'abord, des préoccupations déjà anciennes du projet quant à la génération aléatoire uniforme de configurations combinatoires complexes ont été renouvelées par l'article [6] où interviennent C. Banderier et P. Flajolet. Dans une certaine mesure, l'objectif se relie à la génération de jeux de tests pour la validation de programmes ou de conjectures, certains de ces aspects (algorithmique rapide) étant poursuivis par le projet POLKA de l'UR de Nancy. Ici, ce qui est visé est la génération aléatoire de structures qui n'obéissent pas à des règles de décomposition simple se prêtant à la "méthode récursive" élaborée par P. Flajolet, P. Zimmermann et Bernard Van Cutsem (Université Joseph Fourier, Grenoble) et reprise dans le système MAPLE. Le travail [6] fournit des réponses conséquentes quant à la nature des objets combinatoires qui sont générables par arbres de possibilités et chemins aléatoires. C. Banderier termine actuellement une thèse (soutenance prévue pour mars 2001) concernant des modèles très généraux de chemins (voisins de ce qui est étudié par Guy Fayolle et Vadim Malyshev au projet MEVAL) et de cartes. Les applications se situent largement dans le domaine de la génération aléatoire (arbres de génération) et des propriétés statistiques des graphes planaires ou non [6, 22]. D'autres aspects des rapports entre chemins et graphes sont illustrés par [24, 34], travaux élaborés en commun avec le CTI de Patras, Grèce. Ainsi, certains problèmes de robustesse d'interconnection dans les réseaux à géométrie aléatoire deviennent quantifiables grâce à la combinatoire analytique des chemins. Enfin, l'article [13] est issu d'une collaboration entre le projet et Fabrice Guillemin (France Telecom R&D, Lannion). Il fournit un cadre nouveau à l'étude de certaines files d'attente à fort volume de trafic. Ce travail qui relève *in fine* de l'analyse de réseaux est décrit à la section 6.4. On notera ici qu'il établit un pont entre théorie combinatoire des chemins et de nombreux modèles probabilistes classiques, tout en conduisant à des analyses nouvelles.

Ainsi, à travers une apparente diversité technique, ces recherches participent à l'entreprise globale qui consiste à délimiter précisément le champ des méthodes de la combinatoire analytique. Les objets traités sont des objets de base de l'informatique : arbres, chemins, allocations aléatoires, mots et quantité d'information, hachage, gestion de caches, graphes, configurations géométriques, etc.

Analyse d'algorithmes. Les méthodes d'accès rapide à l'information se classent grosso modo en deux catégories. D'une part, les méthodes fondées sur des comparaisons des objets entre eux (le tri rapide Quicksort, les arbres de recherche, les arbres quadrants, etc.), d'autre part les méthodes opérant par calculs d'adresses (le hachage, les méthodes digitales). Les comportements dans le cas le pire, très étudiés jusqu'à une période récente, ne constituent pas un critère de choix valable, car ils ne détectent que des pathologies dont la probabilité est souvent infime.

En ce qui concerne la recherche par comparaisons, Marianne Durand (ENS) a effectué son stage de DEA sur le thème de l'holonomie et de ses applications, tant en combinatoire analytique qu'en analyse d'algorithmes. L'objet est d'élucider l'expressivité du cadre mathématiquement puissant que sont les fonctions holonomes (voir également les sections relatives

au calcul formel dans ce rapport). Les premières applications concernent les arbres de recherche et plus généralement la recherche par comparaison, ainsi que (tout récemment) l’optimisation combinatoire et le thème de la “complexité paramétrée”. M. Durand aborde actuellement une thèse sous la direction de P. Flajolet. Par ailleurs, un article [18] dont les auteurs sont Hosam Mahmoud (U. de Washington), P. Jacquet (Projet HIPERCOM), M. Régnier et P. Flajolet, met en place tout un ensemble de méthodes pour l’analyse en moyenne et en distribution des principaux algorithmes de tri et de sélection fondées sur les méthodes de calcul d’adresse. Il est ainsi établi qu’existe une algorithmique rapide (par exemple, pour l’extraction de “quantiles” en traitement de données statistiques) valable pour des classes de données naturelles.

En ce qui concerne les méthodes digitales, on retiendra cette année un travail de P. Flajolet et B. Vallée (U. de Caen, membre associé au projet) [27]. Le point de départ est l’analyse de l’algorithme HAKMEM de comparaison de rationnels ou réels par développements en fractions continues. Cet algorithme stable et robuste se relie au calcul formel (arithmétique des systèmes, Projet POLKA) et est développé en géométrie algorithmique (travaux du projet PRISME) ; il se rattache également à la structure de *trie* ou arbre digital qui est l’une des plus fondamentales pour le traitement de données textuelles et fait l’objet de travaux continus dans le projet. Le point de vue qui est adopté relie le problème à la théorie des opérateurs de transfert issue des systèmes dynamiques et du “cadre thermodynamique” (Ruelle). Ce point de vue original introduit en analyse d’algorithmes par B. Vallée est développé conjointement avec J. Clément (doctorant co-encadré, U. de Caen) qui a soutenu récemment une thèse sur ces questions [4]. L’idée consiste très simplement à relier les paramètres d’arbres digitaux à la géométrie d’intervalles fondamentaux (concept issu de la théorie métrique des fractions continues), puis à quantifier par opérateurs les propriétés stochastiques des paramètres fondamentaux déterminant la complexité des algorithmes, grâce à l’analyse d’objets dérivés (séries génératrices, séries de Dirichlet, transformées de Mellin). La méthode est très “versatile” et les nombreuses ramifications en sont poursuivies par B. Vallée et son équipe à l’Université de Caen. Dans [27], il est établi le caractère en moyenne linéaire de l’algorithme HAKMEM, validant ainsi les observations empiriques de Jean-Daniel Boissonnat *et al.* et justifiant l’excellente tenue en moyenne et en probabilité de l’algorithme HAKMEM. Dans le même article, il est montré le fait surprenant que l’algorithme de tri par fraction continue présente un comportement quantitatif précisément équivalent à la fameuse hypothèse de Riemann.

Enfin, différents liens continuent de se tresser entre théorie de l’information et analyse d’algorithmes. L’article [26] est initialement motivé par le cadre de la “théorie analytique de l’information” développé par Wojciech Szpankowski et P. Jacquet (Projet HIPERCOM) dans une longue série de travaux. L’objectif en est d’obtenir des évaluations très précises en théorie de l’information par des méthodes inspirées de la combinatoire analytique. Dans un domaine voisin, une collaboration a commencé avec John Kieffer, spécialiste de la compression de données. Le travail résumé en [28] établit pour la première fois un parallèle de structure entre méthodes de compression dites grammaticales et la structure classique de BDD (*Binary Decision Diagram*) par ailleurs discutée cette année au sein du séminaire ALGO (*cf.* l’exposé de Jean Vuillemin, ENS). Ceci ouvre la voie aux transferts de résultats de théorie de l’information en direction de certains domaines de la complexité des fonctions booléennes (objets d’étude du Projet CODES) et, par exemple, il est mis en évidence une relation entre densité de satisfiabilité, entropie, et taille des BDD résultants.

En résumé ces travaux nous conduisent à une quantification de plus en plus précise des algorithmes et structures de données fondamentales dans des domaines aussi divers que la recherche et la sélection rapides, l'algorithmique "semi-numérique", la conception de circuits, ou le traitement des données textuelles.

6.2 Calcul formel

Participants : Frédéric Chyzak, Marni Mishna, Bruno Salvy.

L'automatisation de l'analyse d'algorithmes a fait un grand progrès cette année, avec l'identification par M. Mishna [35] d'une classe de grammaires attribuées permettant d'exprimer tous les algorithmes analysés automatiquement par le système antérieur $\Lambda\mathcal{R}^{\Omega}$, et fournissant des fonctions génératrices *multivariées*, ce qui permet d'obtenir directement coût moyen, variance, moments d'ordre supérieur, et ouvre la porte à l'analyse automatique des distributions des coûts.

Depuis quelques années, B. Salvy collabore avec l'équipe GAGE de l'École polytechnique, pour expliciter comment des développements récents sur l'algorithmique des *straight-line programs* peuvent être interprétés de manière à produire des implantations efficaces pour des problèmes de nature géométrique dans des systèmes de calcul formel du type de Maple. Ces travaux ont abouti à un nouvel algorithme de résolution de systèmes polynomiaux dont la complexité est la meilleure connue [14, 15]. Une implantation prototype en MAGMA permet de résoudre numériquement des systèmes polynomiaux de grande taille de manière très compétitive par rapport aux meilleurs logiciels disponibles pour les calculs de bases de Gröbner. Les bases de Gröbner interviennent par ailleurs de manière naturelle dans certains problèmes d'origine combinatoire, un joli exemple fait l'objet de [9].

F. Chyzak a poursuivi sa recherche sur la manipulation des suites et des fonctions spéciales par le calcul formel lors de son séjour post-doctoral à l'université de Linz (Autriche). Ce travail a fourni des algorithmes pour le calcul d'équations fonctionnelles linéaires caractérisant la somme ou l'intégrale d'une expression faisant intervenir des suites et fonctions spéciales [8]. L'objectif est maintenant d'améliorer ces algorithmes de façon à assurer la minimalité de l'ordre des équations obtenues [32].

Une collaboration entre F. Chyzak, Peter Paule et Burkhard Zimmermann du RISC, Otmar Scherzer et Armin Schoisswohl de l'*Industrial Mathematics Institute* (autre équipe de l'université de Linz) a donné lieu à une nouvelle construction pour des ondelettes d'un certain type [7]. Une étape cruciale est la détermination sous forme explicite par le calcul formel d'une famille de coefficients de filtrage.

Une autre collaboration avec P. Paule concerne la préparation d'un chapitre sur les méthodes de calcul formel pour les fonctions spéciales, dans le cadre du projet *Digital Library of Mathematical Functions* (DLMF) du *National Institute of Standards and Technology* (NIST, bureau des standards américain). Ce projet ambitieux vise à fournir une nouvelle édition du "*Handbook of Mathematical Functions*", formulaire de référence depuis 1962, qui sera à la fois disponible en version imprimée (environ 1000 pages) et sous forme électronique (CD-rom et site Web). L'achèvement de ce projet est prévu fin 2002.

6.3 Algorithmiques des séquences

Participants : Bruno Boussard, Frédéric Chyzak, Philippe Flajolet, Pierre Nicodème, Mathieu Raffinot, Mireille Régnier, Bruno Salvy, Mathias Vandenbogaert.

Le projet a poursuivi des recherches sur la statistique des mots. Il s'agit de compter le nombre d'occurrences d'un motif donné ou d'un ensemble de motifs dans un texte sous des contraintes variées (distance minimale entre les mots, types de chevauchement autorisés, etc.). Un tel comptage se ramène à l'évaluation des inversions de systèmes d'équations algébriques satisfaits par les séries génératrices.

P. Nicodème, B. Salvy et P. Flajolet ont considéré dans [19] les expressions régulières. Leur algorithme permet de calculer de manière exacte ou asymptotique l'espérance et la variance du nombre d'occurrences de tout motif dans des textes de grande taille dont les lettres obéissent à une distribution de Markov. À part quelques cas de dégénérescence, la distribution limite est gaussienne. L'implantation correspondante a été réalisée par P. Nicodème en Maple, et appliquée aux statistiques d'occurrences de motifs dans les protéines.

Une approche reliant les calculs de moyenne ou de variance à l'étude de langages particuliers est présentée dans [20]. Elle conduit à des formules explicites. Il s'ensuit [31] que la complexité du calcul dépend de la taille minimale d'un automate de "recouvrement". M. Régnier et M. Raffinot ont montré que le nombre total d'états de l'automate de recouvrement est égal au nombre d'états finaux de l'automate de recherche minimal. L'algorithme de calcul de l'automate de recouvrement est l'objet d'un travail en cours, où l'on se base sur une généralisation récente par M. Raffinot et G. Navarro d'un algorithme de recherche d'expressions réégulières.

Il apparaît de plus que le calcul effectif des formules explicites générales [20] se simplifie lorsque les ensembles de mots ont une structure. D'un point de vue formel, on s'intéresse en particulier aux expressions régulières et aux motifs approchés [29] qui apparaissent dans de très nombreux problèmes de biologie moléculaire. Ceci fait l'objet de différentes collaborations. M. Régnier et Alain Denise (Université d'Orsay) travaillent à l'implantation effective des formules, pour différentes structures d'ensembles de mots. B. Boussard a réalisé un début d'implantation en C et les fonctions correspondantes seront rendues accessibles sur la page Web de REMAG (Action de Recherche Coopérative de la direction scientifique de l'INRIA).

Marie-France Sagot et Laurent Marsan (Institut Pasteur) ont présenté à RECOMB'00 un algorithme de reconnaissance de signaux multiples dans les promoteurs. L'évaluation de la pertinence statistique des motifs extraits est faite par *shuffle* de séquences et représente une des phases potentiellement coûteuses de l'algorithme. L'écriture des formules associées a été réalisée. Elles sont simplement calculables et en cours d'implantation par L. Marsan. Leur portée apparaît plus générale, et leur utilisation pour la recherche de signaux de polyadénylation dans des gènes humains est le sujet d'un travail de M. Vandenbogaert en collaboration avec le BGMA (Marseille).

M. Régnier, Alexandre Lifanov and Vsevolod Makeev ont étudié de manière systématique la validité de formules d'approximation. Dans [29], sont justifiées ou corrigées des approximations utilisées dans de nombreux travaux de biologistes (identifiant la moyenne et la variance) et donné des critères simples justifiant l'approximation du modèle markovien par le modèle

Bernoulli. Par ailleurs, la théorie des grandes déviations permet un calcul très précis de probabilités sur des mots, lorsque les calculs exacts sont très coûteux ou numériquement instables. Une première application à la recherche de signaux de polyadénylation a permis de valider cette approche. Enfin, les grandes déviations apparaissent un outil très puissant pour évaluer comment la sur-représentation de certains mots dans une famille de séquences modifie la distribution des autres mots. Une approche utilisant les séries multivariées est l'objet d'une collaboration avec A. Denise.

De nombreux problèmes combinatoires se posent en biologie moléculaire. Dans [16], on étudie une distance particulière, la distance de Levenshtein. Une telle distance peut servir de filtrage pour des recherches de similitudes entre les séquences, qui sont algorithmiquement coûteuses. Dans [30], nous établissons de nouvelles propriétés combinatoires des périodes, puis des quasi-périodes.

6.4 Algorithmique et modélisation des réseaux

Participants : Jean-François Dantzer, Mostapha Haddani, Philippe Robert.

Allocation de bande passante dans un réseau On s'intéresse ici à un modèle d'allocation de ressources : divers types de trafic arrivent avec des demandes de débit très différentes ; il est toujours possible par un mécanisme de réservation et en segmentant ces trafics de leur allouer la totalité de la bande passante. L'inconvénient de ce type de méthode réside dans le traitement ultérieur qui est requis pour recomposer les trafics initiaux. Ces problèmes peuvent se formuler en terme de *bin packing* dynamique : une boîte de taille 1 (la capacité maximale du réseau) et des arrivées de pièces dont la taille est inférieure à 1, chaque pièce demandant à rester dans la boîte le temps de son service (autrement dit dans le cas des réseaux, la requête occupe une fraction de la bande passante). Les questions naturelles qui se posent concernent le débit maximal d'un tel système et la politique d'allocation qui maximise celui-ci. Ce modèle a été étudié, dans un cadre différent, dans un article de Claude Kipnis et P. Robert (1990) ; le débit maximal avait été calculé dans le cas où les temps de service sont exponentiels, les tailles des pièces indépendantes équidistribuées et la discipline de service premier arrivé-premier servi. Les problèmes d'allocation de bande passante ont mis en valeur tout l'intérêt de ce type de modèle, voir les travaux récents de Ed Coffman (Columbia University) et Alexander Stolyar (Lucent Technology).

J.-F. Dantzer et P. Robert étudient actuellement le cas de la discipline *First Fit* : sous les hypothèses probabilistes mentionnées ci-dessus, un message est alloué si la place résiduelle dans la boîte le permet et si aucun des messages arrivés avant lui dans la file d'attente ne peut être alloué. La difficulté de ce modèle vient du fait que la suite des tailles des pièces dans la file d'attente n'est pas une suite indépendante, en raison des prélèvements successifs à l'intérieur de celle-ci. L'espace d'états décrivant le modèle est de dimension dénombrable, donc délicat à manipuler. Les problèmes techniques soulevés sont similaires à ceux, non résolus pour l'instant, des réseaux de files d'attente multi-classe FIFO.

L'étude d'un modèle simple pour lequel il n'y a que deux tailles de messages possibles, 1 et a , a été terminée récemment [10]. Cette année nous avons réalisé l'étude de cet algorithme

dans le cas où il y a trois tailles de pièces possibles. Cet exemple est significativement plus complexe que celui traité précédemment :

- La condition de stabilité de l'algorithme n'est pas la condition "naturelle", de plus son expression fait intervenir une fonctionnelle quadratique des paramètres du modèle ;
- La preuve de la stabilité utilise le cadre des limites fluides. Les limites fluides n'existent pas d'emblée sur tous les états comme c'est le cas habituellement. Pour décrire convenablement le système dynamique associé, on introduit la notion d'état initial "lisse". Il faut ensuite montrer que le processus de Markov partant d'un état initial arbitraire atteint un tel état dans un temps de l'ordre de la taille de l'état. Cette notion d'état lisse devrait permettre, nous semble-t-il, d'attaquer certains problèmes de stabilité de réseaux multi-classe non résolus pour l'instant ;
- Le système dynamique obtenu se ramène à un produit de matrices aléatoires de tailles 2×2 . Son exposant de Liapounov est calculé.

Optimisation de la gestion du trafic TCP dans un réseau Une proposition d'étude de l'allocation de bande passante dans les réseaux a été acceptée dans le cadre des consultations thématiques de France Telecom R&D. Cette étude se fait en collaboration avec Alain Dupuis et Fabrice Guillemin (Lannion).

Elle se propose d'étudier les problèmes d'allocation de bande passante dans des réseaux transportant du trafic élastique, comme par exemple le trafic TCP dans le réseau INTERNET. Actuellement dans un réseau de télécommunications les messages se partagent la bande passante de façon égalitaire. Par exemple, l'implémentation de TCP est telle que les ajustements se font sur les nœuds les plus chargés, à ces nœuds la bande passante est équitablement répartie entre les messages. Si les mécanismes de ce type de politique ont l'avantage de réguler correctement, de façon distribuée le trafic, ils présentent l'inconvénient de ne pas utiliser pleinement la capacité du réseau. En effet, en prenant un cas simplifié, si un message passe par une série de N nœuds peu chargés sauf un où passent M messages, en supposant que la bande passante maximum soit λ à chaque nœud, le protocole TCP fera que le message sera transmis au taux λ/M à travers le réseau. Globalement une petite fraction λ/M de la capacité totale du réseau sera utilisée par nœud au lieu de λ dans le cas idéal.

Il est donc intéressant d'étudier la possibilité d'augmenter l'utilisation de la capacité d'un réseau en modifiant les algorithmes de partage de bande passante à chaque nœud tout en préservant certaines des caractéristiques actuelles du protocole : les mécanismes doivent être distribués et adaptatifs. C'est le but principal de l'étude menée avec France Telecom R&D.

Dans un premier temps nous étudions une politique de traitement de partage égalitaire (Processor-sharing). Elle est centrale dans les implémentations actuelles de TCP. Plusieurs types de renormalisation sont actuellement analysés (lois des grands nombres, théorème de la limite centrale et "heavy traffic").

Divers Jean-François Dantzer a soutenu sa thèse "Stabilité des réseaux de files d'attente et limites fluides stochastiques" le 28 janvier 2000 devant le jury constitué de F. Charlot (Univer-

sité de Rouen), B. Chauvin (Université de Versailles), P. Flajolet, A. Mokkadem (Université de Versailles), B. Ycart (Université de Paris V), et P. Robert.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Algorithmique et modélisation des réseaux

Mostafa Haddani et *Philippe Robert* participent à la consultation thématique de France Telecom R&D, sur l'optimisation de la gestion du trafic TCP dans un réseau (voir la section 6.4). Ce contrat est d'une durée de deux ans.

7.2 Calcul formel

Le projet ALGO et la compagnie *Waterloo Maple Inc.* ont développé une collaboration très étroite fondée sur des intérêts réciproques. D'une part il est intéressant pour la compagnie d'intégrer des fonctionnalités à la pointe de la recherche en calcul formel (voir la section 3.2). D'autre part cette intégration fournit aux programmes réalisés par les membres du projet un grand nombre d'utilisateurs d'origines très diverses. Cette relation étroite nous permet également de participer aux choix effectués par les développeurs du système.

De nombreux échanges ont ainsi lieu entre le projet et la compagnie. J. Carette est retourné à la compagnie WMI, après une participation de plus de trois ans au projet ALGO. Il y est maintenant *Product Development Director* et s'occupe de tous les aspects techniques de Maple. De même, E. Murray, après avoir passé plus de deux ans au projet ALGO à programmer le *package* COMBSTRUCT de Maple, travaille maintenant à la compagnie WMI.

L'arrivée de Maple dans l'enseignement en classes préparatoires a nécessité un important travail de formation des enseignants. Les membres du projet ont participé activement à cet effort. Claude Gomez (projet), B. Salvy et P. Zimmermann (projet POLKA) ont écrit un livre il y a quatre ans sur l'utilisation de Maple. Depuis deux ans, P. Dumas consacre un effort important à la rédaction des solutions des exercices de ce livre et à leur mise à la disposition de tous sur le Web.

Grâce à ces nombreuses activités autour de Maple, la compagnie WMI considère l'INRIA comme un partenaire privilégié et lui accorde une licence site gratuite couvrant l'ensemble des centres. Une quinzaine de projets utilisent ce système à des degrés divers.

Un club des utilisateurs francophones de Maple a été créé il y a trois ans. La liste de diffusion et le site Web de ce club sont hébergés par le projet ALGO, et gérés en commun avec Gilles Villard et Claire Di Crescenzo de l'IMAG.

8 Actions régionales, nationales et internationales

8.1 Actions nationales

En 2000 s'est poursuivie une Action de Recherche Coopérative de la direction scientifique de l'INRIA intitulée ALCOPHYS. Les maîtres d'œuvre en sont le LABRI de Bordeaux (Robert Cori) et le Projet Algorithmes, et y participent une équipe du LRI (Claire Kenyon) et une équipe de

physiciens de l'Université Paris-Sud. Le thème en est le rapprochement et la confrontation entre méthodes probabilistes, combinatoire analytique et physique statistique dans le but d'analyser les systèmes complexes de grande taille.

Le Projet ALGO participe à un contrat entre Elf-Sanofi et le Projet PRISME de l'INRIA-Sophia. L'une des études de ce contrat (signé en l'an 2000) poursuit des travaux lancés par Frédéric Cazals (en 1997–1998) lors de son séjour post-doctoral au sein du projet ALGO. Il s'agit de la recherche de similarités moléculaires pour laquelle une algorithmique randomisée et des structures digitales voisines des *skip-lists* conduisent à des algorithmes de classification multidimensionnelle très performants.

M. Régnier participe à l'Action de Recherche Coopérative de la direction scientifique de l'INRIA intitulée REMAG pilotée par le LORIA (G. Kucherov) qui rassemble des chercheurs de l'IRISA, de l'Institut Pasteur, des Universités de Bordeaux et d'Orsay. L'objectif est la recherche de motifs dans le génome.

Aléa est un groupe de travail dédié à l'analyse d'algorithmes et à l'analyse des propriétés des structures aléatoires discrètes. Il s'agit d'un pôle de rencontre entre informaticiens et probabilistes travaillant dans le domaine des modèles discrets. L'activité est soutenue actuellement par le groupement A.L.P. (qui encapsule l'ancien GDR/PRC A.M.I.) et est animée par P. Flajolet. Un atelier rassemblant une quarantaine de participants s'est tenu à Marseille en mars 2000.

8.2 Actions financées par la Commission Européenne

Le projet ALGO est pour une période de trois ans 2000–2003 l'une des composantes du projet ESPRIT *Long Term Research* ALCOM-FT. Ce projet rassemble dix groupes leaders dans le domaine de la recherche algorithmique en Europe. L'objectif affiché est la découverte de nouveaux concepts algorithmiques et l'identification des algorithmes clefs transverses à de nombreuses applications. Quatre directions de travail ont été identifiées : *(i)* ensembles de données massifs ; *(ii)* systèmes de communication complexes ; *(iii)* optimisation en production et planification ; *(iv)* recherches méthodologiques et expérimentales en algorithmique. Les travaux du projet se situent principalement dans les axes *(ii)* et *(iv)*.

Le projet pilote la composante française du projet INTAS 99-1476, *Methods, algorithms and software for functional and structural annotation of complete genomes*, d'une durée de dix-huit mois. Ce projet implique quatre autres équipes (russe, allemandes (EMBL et MIPS) et autrichienne).

8.3 Relations bilatérales internationales

Une collaboration entre l'Engelhardt Institute of Computational Biology (Moscou) et le projet ALGO est soutenue depuis juin 2000 par l'Institut Liapounov.

8.4 Accueils de chercheurs étrangers

Une grande partie de nos invités ont fait des exposés au séminaire du projet. Cette année, nous avons accueilli : Wojciech Szpankowski, Purdue University, West Lafayette, USA ; Ilan Vardi, visiteur à l'IHES ; François Bergeron du LaCIM, Université du Québec à Montréal ;

Mamoru Hoshi, University of Electro-Communications, Japon, du 9 au 31 octobre ; Ralph Neininger, Institut für Mathematische Stochastik, Freiburg, Germany, du 13 au 18 novembre ; John Shackell et James Beaumont, University of Kent at Canterbury, UK, du 7 au 11 février ; Peter Paule, RISC, Linz, Austria du 26 septembre au 3 octobre.

Marni Mishna, Simon Fraser University, British Columbia, Canada, a séjourné plusieurs mois au sein du projet, du 15 mars au 30 avril et du 19 juin au 31 juillet ; Robert Sedgewick, Princeton University, USA a passé le mois de mai à l'Inria et il a à cette occasion présenté un exposé le 16 mai au Colloquium de Rocquencourt. Jason Gao, Carleton University, Canada, est venu du 3 au 20 juin.

Dans le cadre ALCOPHYS, nous avons eu la visite de Elchanan Mossel, Hebrew University of Jerusalem, Israël, du 1er septembre 1999 au 31 mars 2000 et le LaBRI, Bordeaux, a reçu William Orrick, University of Melbourne, Australia, du 1er mars au 31 août.

Dans le cadre de l'action REMAG, Vsevolod Makeev, Engel'hardt Institute of Molecular Biology, Moscow, nous a rendu deux visites, du 9 au 11 mai puis du 9 au 12 octobre.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

Le projet ALGO a un séminaire régulier auquel participent plusieurs équipes partenaires de la région parisienne. Les actes en sont édités et publiés chaque année [1] et [3]. Le workshop pluridisciplinaire ALÉA déjà évoqué a rassemblé une quarantaine de chercheurs. Enfin, le projet joue un rôle très important dans l'animation de la communauté internationale travaillant en Analyse d'Algorithmes : P. Flajolet préside le Comité de Pilotage de la série des séminaires "*Analysis of Algorithms*" et en juillet 2000, s'est tenu à Krinica Morska (Pologne) le "*Sixth Seminar on Analysis of Algorithms*" avec environ 70 participants dont une forte présence française.

Cyril Banderier est l'un des responsables de l'Association des Doctorants de l'INRIA Rocquencourt, il anime pour la seconde année le Colloquium Junior de l'INRIA Rocquencourt, et a fait une intervention le 7 mars 2000, aux Mureaux, dans le cadre de la manifestation "Jours de Sciences".

Philippe Flajolet a été membre du comité de programme du "*Sixth Seminar on Analysis of Algorithms*" (Pologne), est directeur du groupe de travail ALÉA, membre du comité de programme de la conférence "*Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*" (Versailles, Septembre 2000). Avec B. Salvy, il a animé une journée à l'École des Jeunes Chercheurs en algorithmique (Caen, mars 2000). P. Flajolet est éditeur de la revue *Random Structures and Algorithms* (Wiley) et "*Honorary Editor*" de *Theoretical Computer Science* (Elsevier). Il a été membre de jury et souvent rapporteur (*) des thèses (*T*) ou habilitations (*H*) suivantes : Hoang Ngoc Minh (*H*, Lille, calcul formel), J.-F. Dantzer (*T*, Versailles, probabilités), S. Corteel (*T*, Orsay, combinatoire), C. Zhang (*H**, La Rochelle, analyse complexe), P. Gaudry (*T*, Polytechnique, cryptographie), J. Jabbour (*T**, Versailles, probabilités), V. Ravelomanana (*T**, Amiens, combinatoire), C. Nicaud (*T**, Paris 6, automates).

Mostafa Haddani et *Philippe Robert* participent à la consultation thématique de France Telecom R.&D. sur l'optimisation de la gestion du trafic TCP dans un réseau. Ce contrat est d'une durée de deux ans.

Mireille Régnier a été rapporteur de la thèse de Sabine Mercier (Université de Rouen) et a organisé la venue d'Edward Trifonov (Weizmann Institute of Science, Israël) au Colloquium de Rocquencourt.

Bruno Salvy est membre du comité éditorial du *Journal of Symbolic Computation*.

9.2 Enseignement universitaire

Cyril Banderier, moniteur à l'université de Paris-Nord (P13) enseigne en premier cycle universitaire (première et deuxième année du Deug, options maths-info et ingénieur) différentes matières informatiques.

Jean-Francois Dantzer enseigne en licence de mathématiques à l'université Versailles Saint-Quentin.

Marianne Durand vacataire à l'École polytechnique enseigne en tronc commun d'informatique.

Philippe Flajolet enseigne avec B. Vallée un cours "Modèles Combinatoires" au sein du DEA Algorithmique de la région parisienne.

Mostafa Haddani enseigne en DEUG STPI et en DEUG Économie à l'université Versailles Saint-Quentin.

Philippe Robert a donné un cours de DEA intitulé "Théorèmes limites pour l'étude des files d'attente" au Laboratoire de Probabilités de l'Université de Paris VI. Il enseigne le cours "Processus stochastiques" au DEA Maths-Info de l'Université de Versailles Saint-Quentin (avec A. Rouault et J.-M. Fourneau).

9.3 Participation à des colloques, séminaires, invitations

Cyril Banderier a assisté et donné des exposés aux Journées ALÉA à Marseille, à l'École Jeunes Chercheurs de Caen, à la conférence internationale Séries Formelles et Combinatoire Algébrique (SFCA/FPSAC) à Moscou, au séminaire AOFA à Gdansk, ainsi qu'au Séminaire Algorithmes (INRIA Rocquencourt). Il a par ailleurs assisté au Colloque en l'honneur de Michel Mendès-France (Bordeaux) et au Colloque Informatique et Mathématiques à Versailles.

Frédéric Chyzak a présenté ses résultats sur la manipulation de suites et fonctions spéciales par le calcul formel dans le cadre d'un workshop spécialisé du MSRI (*Computational Algebraic Analysis*, Berkeley, USA) et d'une invitation au grand colloque généraliste de SIAM sur les mathématiques discrètes (*Tenth SIAM Conference on Discrete Mathematics*, Minneapolis, USA). Chacune de ces deux fois, l'exposé théorique a été complété d'une démonstration logicielle de ses programmes, et tout particulièrement d'MGFUN.

Marianne Durand a assisté à l'École Jeunes Chercheurs de Caen, aux journées ALEA 2000 à Marseille, au séminaire AofA à Gdansk (Pologne) et au Colloque Informatique et Mathématiques à Versailles.

Philippe Flajolet a été conférencier invité à la conférence "*Mathematics and its Role in Civilization*" (Macau, janvier 2000) ainsi qu'à la conférence LATIN 2000 (Montevideo, avril 2000) et au "*Colloquium on Mathematics and Computer Science: Algorithms, Trees, Combinatorics and Probabilities*" (Versailles, septembre 2000). Il a présenté diverses communications (invitées ou non) aux journées ALÉA (Marseille, mars 2000), au colloque ICALP (Genève, juillet 2000), au *Workshop on Experimental Algorithmics* (Dagstuhl, septembre 2000), aux journées Méthodes aléatoires et statistiques de la SMAI (Rennes, septembre 2000), Il a également été invité à l'université de Purdue pour une collaboration avec W. Szpankowski sur le thème de la théorie analytique de l'information.

Mireille Régnier a participé à la conférence RECOMB'00 à Tokyo, et présenté une communication aux conférences JOBIM'00 (Montpellier), CPM'00 (Montréal), GCB'00 (Heidelberg) ainsi qu'aux journées MAS (Méthodes aléatoires et statistiques) de la SMAI. Elle a fait des séminaires au LIX-LRI, à Marseille et Lille, à l'Institut de Biologie Moléculaire Engelhardt (Moscou), à l'Université de Chofu (Tokyo) et au centre d'IBM-Yorktown (USA).

Philippe Robert a participé à la conférence INFORMS à Boca Raton (USA) pour présenter les travaux sur les problèmes d'allocation de bande passante [10]. Il a rendu visite à Isi Mitrani (Université de Newcastle) en septembre. Il a en outre fait un exposé sur les files d'attente multi-classe au séminaire de l'Université de Versailles-St Quentin, de l'Université Henri Poincaré à Nancy et du laboratoire de mathématiques appliquées de l'École polytechnique.

Bruno Salvy a été invité à présenter un exposé sur son nouvel algorithme de calcul d'intégrales définies à un *Workshop on Computational Algebraic Analysis* du MSRI (Berkeley) et à un *Symposium on Asymptotics and Applied Analysis* de San Diego (Californie). Il a donné un exposé sur le même thème aux *Journées Nationales de Calcul Formel* à Aussois. Il a aussi été l'un des conférenciers invités principaux de la *SIAM Conference on Discrete Mathematics* où il a présenté les activités du projet relatives au calcul formel pour l'analyse combinatoire et l'analyse d'algorithmes. Il a par ailleurs été invité au *National Institute of Standards and Technology* (Washington) où il a présenté les développements récents de l'algorithmique des fonctions holonomes et la manière dont ceux-ci pourraient s'intégrer dans le développement de la *Digital Library of Mathematical Functions*. Enfin, il a fait un mini-cours d'analyse d'algorithmes à l'École des Jeunes Chercheurs en Algorithmique (Caen, Mars 2000).

10 Bibliographie

Livres et monographies

- [1] F. CHYZAK (éditeur), *Algorithms Seminar, 1999–2000, Research Report, 4056*. Institut National de Recherche en Informatique et en Automatique, novembre 2000.
- [2] P. ROBERT, *Réseaux et files d'attente: méthodes probabilistes, Mathématiques et Applications, 35*, Springer-Verlag, Berlin, Octobre 2000.
- [3] B. SALVY (éditeur), *Algorithms Seminar, 1998–1999, Research Report, 3830*. Institut National de Recherche en Informatique et en Automatique, décembre 1999.

Thèses et habilitations à diriger des recherches

- [4] J. CLÉMENT, *Arbres Digitaux et Sources Dynamiques*, Thèse de doctorat, Université de Caen, septembre 2000.
- [5] J.-F. DANTZER, *Stabilité des réseaux de files d'attente et limites fluides stochastiques*, thèse de doctorat, Université de Versailles-St Quentin, 2000.

Articles et chapitres de livre

- [6] C. BANDERIER, M. BOUSQUET-MÉLOU, A. DENISE, P. FLAJOLET, D. GARDY, D. GOUYOU-BEAUCHAMPS, «Generating Functions for Generating Trees», *Discrete Mathematics*, 2000, (A short version "On Generating Functions of Generating Trees" appeared in the proceedings of FP-SAC'99 and as INRIA Report 3661), <ftp://algo.inria.fr/banderier/Papers/DiscMath99.ps>.
- [7] F. CHYZAK, P. PAULE, O. SCHERZER, A. SCHOISSWOHL, B. ZIMMERMANN, «The construction of orthonormal wavelets using symbolic methods and a matrix analytical approach for wavelets on the interval», *Experimental Mathematics*, 2000.
- [8] F. CHYZAK, «An extension of Zeilberger's fast algorithm to general holonomic functions», *Discrete Mathematics* 217, 1-3, 2000, p. 115–134, Formal power series and algebraic combinatorics (Vienna, 1997).
- [9] R. CORI, D. ROSSIN, B. SALVY, «Polynomial Ideals for Sandpiles and their Gröbner Bases», *Theoretical Computer Science*, 2000, à paraître. Version préliminaire dans le rapport de recherche INRIA n^o 3946, <http://www.inria.fr/rrrt/rr-3946.html>.
- [10] J.-F. DANTZER, M. HADDANI, P. ROBERT, «On the stability of a bandwidth packing algorithm», *Probability in the Engineering and Informational Sciences* 14, 1, 2000, p. 57–79.
- [11] J.-F. DANTZER, I. MITRANI, P. ROBERT, «Large Scale and Heavy Traffic Asymptotics for Systems with Unreliable Servers», *Queueing Systems, Theory and Applications*, 2000, À paraître.
- [12] V. DUMAS, P. ROBERT, «On the maximal throughput of a resource sharing model», *Mathematics of Operation Research*, 2000, À paraître.
- [13] P. FLAJOLET, F. GUILLEMIN, «The Formal Theory of Birth-and-Death Processes, Lattice Path Combinatorics, and Continued Fractions», *Advances in Applied Probability* 32, 2000, p. 750–778.

- [14] M. GIUSTI, K. HÄGELE, G. LECERF, J. MARCHAND, B. SALVY, «Computing the Dimension of a Projective Variety: the Projective Noether Maple Package», *Journal of Symbolic Computation* 30, 3, septembre 2000, p. 291–307, Version préliminaire dans le rapport de recherche INRIA n^o 3224, <http://www.inria.fr/rrrt/rr-3224.html>.
- [15] M. GIUSTI, G. LECERF, B. SALVY, «A Gröbner Free Alternative for Polynomial System Solving», *Journal of Complexity*, 2000, 51 pages. À paraître.
- [16] D. HIRSCHBERG, M. RÉGNIER, «Tight Bounds on the Number of String Subsequences», *Journal of Discrete Algorithms*, 2000, À paraître, version préliminaire présentée à CPM'99.
- [17] S. KOTZ, H. MAHMOUD, P. ROBERT, «On Generalized Pólya Urn Models», *Statistics & Probability Letters* 49, 2, 2000, p. 163–173.
- [18] H. MAHMOUD, P. FLAJOLET, P. JACQUET, M. RÉGNIER, «Analytic variations on bucket selection and sorting», *Acta Informatica* 36, 9-10, 2000, p. 735–760.
- [19] P. NICODÈME, B. SALVY, P. FLAJOLET, «Motif Statistics», *Theoretical Computer Science*, 2000, à paraître. Extended version of an article published in the proceedings of 7th Annual European Symposium on Algorithms ESA'99, Prague, July 1999.
- [20] M. RÉGNIER, «A Unified Approach to Word Occurrences Probabilities», *Discrete Applied Mathematics* 104, 1, 2000, p. 259–280, Special issue on Computational Biology.

Communications à des congrès, colloques, etc.

- [21] C. ADJIH, P. JACQUET, P. ROBERT, «Differentiated admission control in large networks», in : *INFOCOM 2000*, IEEE, p. 1455–1461, Tel Aviv, 2000.
- [22] C. BANDERIER, R. P. DOBROW, «A Generalized Cover Time for Random Walks on Graphs», in : *FPSAC'00*, Springer, p. 113–124, juin 2000, <ftp://algo.inria.fr/banderier/Papers/Dobrow/walksgraphs.ps>.
- [23] C. BANDERIER, P. FLAJOLET, G. SCHAEFFER, M. SORIA, «Planar Maps and Airy Phenomena», in : *ICALP'00, Lecture Notes in Computer Science, 1853*, Springer-Verlag, p. 388–402, janvier 2000, <ftp://algo.inria.fr/banderier/Papers/Airy/coal.ps>.
- [24] P. FLAJOLET, K. HATZIS, S. NIKOLETSEAS, P. SPIRAKIS, «Trade-offs between density and robustness in random interconnection graphs», in : *IFIP International Conference on Theoretical Computer Science*, J. van Leeuwen, O. Watanabe, M. Hagiya, P. D. Moses, T. Ito (éditeurs), *Lecture Notes in Computer Science, 1872*, p. 152–168, août 2000. (Proceedings of IFIP TCS'2000, Sendai, August 2000.).
- [25] P. FLAJOLET, M. NOY, «Analytic Combinatorics of Chord Diagrams», in : *Formal Power Series and Algebraic Combinatorics*, D. Krob, A. A. Mikhalev, A. V. Mikhalev (éditeurs), p. 191–201, 2000. (Proceedings of the 12th International Conference, FPSAC'2000; June 2000, Moscow.).
- [26] P. FLAJOLET, W. SZPANKOWSKI, «Analytic Variations on the Redundancy Rate of Renewal Processes», in : *2000 IEEE International Symposium on Information Theory*, IEEE Information Theory Society, p. 499, 2000. (Short abstract).

- [27] P. FLAJOLET, B. VALLÉE, « Continued Fractions, Comparison Algorithms, and Fine Structure Constants », in : *Constructive, Experimental, and Nonlinear Analysis*, M. Théra (éditeur), *Canadian Mathematical Society Conference Proceedings, 27*, American Mathematical Society, p. 53–82, Providence, 2000.
- [28] J. KIEFFER, P. FLAJOLET, E.-H. YANG, « Data Compression Via Binary Decision Diagrams », in : *2000 IEEE International Symposium on Information Theory*, IEEE Information Theory Society, p. 296, 2000. (Short abstract).
- [29] M. RÉGNIER, A. LIFANOV, V. MAKEEV, « Three variations on word counting », in : *GCB'00*, Logos-Verlag, p. 75–82, 2000. Proc. German Conference on Bioinformatics, Heidelberg.
- [30] M. RÉGNIER, L. MOUCHAR, « Periods and quasiperiods characterization », in : *CPM'00, Lecture Notes in Computer Science, 1848*, Springer-Verlag, p. 388–396, 2000. Proc. 11-th Annual Symposium on Combinatorial Pattern Matching, CPM'00, Montreal.
- [31] M. RÉGNIER, « Complexity of Unusual Words Counting », in : *JOBIM'00, Lecture Notes in Computer Science*, Springer-Verlag, 2000. à paraître, Proc. of JOBIM'00, Montpellier.

Rapports de recherche et publications internes

- [32] F. CHYZAK, « About the non-minimality of the outputs of Zeilberger's algorithm », *Numerical and Symbolic Scientific Computing, Research Report n° 00-08*, Austrian project SFB F013, Linz, Austria, avril 2000, Bruno Buchberger and Peter Paule, Eds. 20 pages.
- [33] J.-F. DANTZER, P. ROBERT, « Analysis of a multi-class queueing system », *Rapport de Recherche n° 4037*, INRIA, octobre 2000, <http://www.inria.fr/rrrt/rr-4037.html>.
- [34] P. FLAJOLET, K. HATZIS, S. NIKOLETSEAS, P. SPIRAKIS, « On the Robustness of Interconnections in Random Graphs: A Symbolic Approach », *Research Report n° 4069*, Institut National de Recherche en Informatique et en Automatique, novembre 2000, 21 pages. à paraître in *Theoretical Computer Science*, 2001.
- [35] M. MISHNA, « Attribute grammars and automatic complexity analysis », *Rapport de Recherche n° 4021*, Institut National de Recherche en Informatique et en Automatique, octobre 2000, <http://www.inria.fr/rrrt/rr-4021.html>.

Divers

- [36] P. FLAJOLET, G. LOUCHAR, « Analytic Variations on the Airy Distribution », Accepted for publication in *Algorithmica*, Special Issue on Analysis of Algorithms 2001, mai 2000, 16 pages.