



Institut des sciences informatiques et de leurs interactions (INS2I - principal)
Institut des sciences de l'ingénierie et des systèmes (INSIS - secondaire)



*Groupe de Recherche en Informatique, Image,
Automatique et Instrumentation de Caen*

Greyc - Umr6072

Projet du GREYC
Évaluation Quadriennale Novembre 2010

Table des matières

Projet du GREYC	1
0.1 Plateforme sécurité	10
0.2 Plateforme e-santé	11
0.3 Plateforme mécatronique et systèmes nomades	11
0.4 Plateforme document numérique	12
1 Équipe Algorithmique, Modèles de calcul, Aléa, Cryptographie, Complexité	14
1.1 L'équipe	15
1.1.1 Les trois thèmes.	15
1.1.2 Les membres.	15
1.1.3 Le contexte de la création de l'équipe AMACC.	16
1.1.4 La politique scientifique de l'équipe AMACC.	17
1.1.5 Risques et opportunités.	18
1.2 Thème 1 : Modèles de calcul, logique et complexité.	18
1.3 Thème 2 : Protection de l'information, codage et cryptographie.	20
1.4 Thème 3 : Structures aléatoires et analyse d'algorithmes.	22
1.5 Les projets de l'équipe en cours en 2011.	24
2 Équipe CONstraints, DATA mining and Graphs	28
2.1 Introduction	29
2.1.1 Cœurs de métiers	31
2.1.2 Collaborations existantes	31
2.2 Projet de recherche	32
2.2.1 Thématiques transversales	33
2.2.2 Projets propres aux "Cœurs de métiers"	35
2.3 Vie de l'équipe	36
3 Équipe Modèles, Agents, Décision	38
3.1 Équipe MAD	39
3.1.1 Composition de l'équipe	39
3.1.2 Perspectives scientifiques	40
3.1.3 Collaborations avec d'autres équipes et disciplines	43
4 Équipe Document numérique, langue, usages	45
4.0.4 Composition probable au 1er janvier 2012	46
4.0.5 Présentation	46
4.0.6 Auto-évaluation	47
4.0.7 Perspectives scientifiques	48
4.0.8 Organisation	51
5 Équipe Monétique & biométrie	52
5.1 Equipe Monétique & Biométrie	53
5.1.1 Composition	53
5.1.2 Auto-évaluation	54
5.1.3 Perspectives scientifiques	55

6	Équipe Image	58
6.1	Perspectives scientifiques	59
6.2	Projet 1 : Variationnel et multiéchelle sur les variétés	59
6.3	Projet 2 : Modélisation géométrique et topologique	60
6.3.1	Motivation	60
6.3.2	Verrous	60
6.3.3	Démarche	61
6.4	Projet 3 : Reconnaissance d'objets dans des images et des vidéos, indexation, classification.	61
6.5	Projet 4 : Ingénierie des connaissances pour la conception d'applications et l'évaluation en imagerie	62
6.5.1	Motivations	62
6.5.2	Verrous	62
6.5.3	Démarche	63
6.6	Vie de l'équipe	63
6.7	Risques	63
6.8	Opportunités	64
6.9	Politique d'intégration de non producteurs	64
6.10	Références bibliographiques	64
7	Équipe Électronique	66
7.1	Equipe Electronique	67
7.1.1	Auto-évaluation	67
7.1.2	Perspectives scientifiques	68
8	Équipe Automatique	72
8.1	Membres permanents	73
8.1.1	Auto-évaluation	73
8.1.2	Perspectives scientifiques	74
8.2	Nouvelle dynamique de recherche au sein de l'équipe	76

Projet du GREYC

Le Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen (GREYC) est une unité de recherche associée au CNRS sous la forme d'une UMR rattachée à titre principal à l'INS2I¹ et à titre secondaire à l'INSIS². Par sa taille (plus de 200 membres permanents ou non permanents) et la qualité de ses travaux, le GREYC est un acteur majeur de la recherche en Normandie. Ses équipes jouent également un rôle important dans le paysage scientifique français et certains travaux bénéficient d'une reconnaissance internationale.

Le GREYC effectue des recherches fondamentales, méthodologiques et appliquées sur les problématiques des STIC autour de l'information et des systèmes. Il développe une approche système issue du potentiel des théories de traitement de l'information et des systèmes pour le développement de modèles, d'algorithmes et d'approches corroborés par des réalisations matérielles et logicielles et des validations expérimentales. Il se distingue aussi par une composante pluridisciplinaire avec des actions interdisciplinaires aussi bien dans le domaine des SHS que dans le domaine des interactions de l'informatique avec les mathématiques, l'EEA et la santé.

Le GREYC a beaucoup évolué au cours du dernier quadriennal : il a su préserver et développer sa dynamique interne, maintenir une cohérence scientifique entre les thèmes de recherche de ses équipes, conduire des projets de recherche innovants, asseoir une meilleure ambiance de travail et d'échanges scientifiques, développer des compétences en matière d'ingénierie des systèmes et favoriser l'émergence de nouveaux projets de recherche entre équipes.

Concernant les équipes, quatre nouvelles équipes sont en cours de formation par reconfiguration des équipes existantes. Tout d'abord, l'équipe « Algorithmique, Modèles de calcul, Aléa, Cryptographie, Complexité » (AMACC) s'est constituée autour des thèmes algorithmique et complexité de l'ancienne équipe « Algorithmique ». Elle développe des travaux en rapport avec la protection de l'information, notamment en codage et cryptographie avec une ouverture sur d'autres thématiques de la sécurité, comme l'internet des objets ou l'environnement embarqué sécurisé. Ensuite, la nouvelle équipe « COnstraints, DAta mining and Graphs » (CoDag) est issue de la fusion d'une partie de l'ancienne équipe « Algorithmique » et d'une partie de l'ancienne équipe « Documents Données Langues ». Elle développera des recherches à la frontière entre la fouille de données et l'optimisation en vue de mettre en évidence les compétences verticales du laboratoire en matière d'aide à la décision. La nouvelle équipe « Document numérique, Langue, Usages » (DLU) est quant à elle constituée pour partie des membres de l'ancienne équipe « Interaction Sémiotique : langues, Diagrammes » et d'une partie de l'équipe « Documents Données Langues ». Elle regroupe ainsi l'ensemble des compétences de l'unité relatives au document textuel, intègre des travaux sur les modalités texte-image et développe des coopérations autour des divers aspects du document numérique (génération, annotation et analyse de collections annotées ou non, usages). Enfin, la quatrième nouvelle équipe, l'équipe « Monétique et Biométrie », résulte elle d'une création, sur la base de l'ancien groupe de même nom. La création de cette équipe est justifiée par l'existence d'une activité de recherche importante sur la sécurisation de transactions électroniques, qui prend en compte la dimension théorique mais aussi usage de ces nouvelles technologies, notamment par une coopération active avec des chercheurs en psychologie cognitive.

Si le dernier quadriennal a vu beaucoup évoluer le paysage de l'enseignement supérieur et de la recherche, que ce soit sur le plan régional ou national, le prochain quadriennal sera probablement tout aussi riche en évolutions, typiquement avec la mise en place des compétences élargies de nos établissements (UCBN et ENSICAEN).

1. Institut des Sciences Informatiques et de leurs Interactions
2. Institut des Sciences de l'Ingénierie et des Systèmes

Sur le plan local, le GREYC doit anticiper les évolutions de l'enseignement supérieur et de la recherche pour maintenir son identité scientifique tout en confortant et préservant la dynamique interne de la communauté STIC en Normandie. Il a été fortement concerné par les restructurations de la recherche liées au projet de création du PRES Normand (Normandie Université). Le GREYC travaille en ce moment sur le montage d'une fédération informatique, baptisée NOMASTIC, qui permettrait une meilleure visibilité nationale des STIC en Normandie. Cette dynamique interrégionale se retrouve également dans un projet d'école doctorale Normande en STIC (école inter-régionale). L'avenir du GREYC est aussi lié aux évolutions des formations régionales, comme la restructuration des spécialités de l'ENSICAEN, la création éventuelle d'une INSA Normande ou encore le projet d'école d'ingénieur ESIX de l'UCBN. Toujours dans cet esprit, un groupement EEA Normand a été constitué à partir d'un projet de Master commun. Les enseignants du GREYC sont conscients de la nécessité d'avoir des formations de haut niveau pour asseoir la recherche (attractivité des établissements, ouvertures de postes, etc.). Notons toutefois que ces projets peinent à se réaliser, que ce soit en raison de difficultés liées à l'éloignement ou des divergences dans les stratégies des établissements.

D'un point de vue national, les effets de la LOLF, de la loi LRU et de la création de l'ANR vont continuer d'avoir un impact fort sur notre laboratoire durant le prochain quadriennal. Ces dispositifs, en imposant à tous les niveaux des évaluations du laboratoire, ont un effet globalement positif et le poussent vers l'excellence (excellence des recherches, des collaborations, des stratégies de valorisation, etc.). Néanmoins, la pression qu'exerce l'évaluation peut avoir des aspects négatifs : certains domaines, e.g. les domaines pluridisciplinaires, dans lesquels il est plus difficile de publier risquent d'être abandonnés ; des recherches qui sont en marge des courants dominants peuvent avoir du mal à subsister ; certains colloques ou revues francophones, pourtant de qualité, seraient délaissés car non pris en compte dans les évaluations, ce qui est regrettable. Nous devons trouver en ce domaine un juste équilibre. L'accroissement des ressources provenant de l'ANR est d'une manière générale très bénéfique pour notre unité. Nous devons cependant veiller à ce que cela ne conduise pas à la morceler et que la cohérence des directions scientifiques soit préservée.

Dans le cadre du volet Equipements d'Excellence, le GREYC est le partenaire STIC d'un projet STIC/SHS (Numérique Normandie) porté par l'Université de Caen et sa MRS³. Numérique Normandie est un projet ambitieux fondé sur les acquis de recherche et le gisement culturel et patrimonial de la région, il trouve sa valorisation économique dans de nouveaux débouchés de diffusion par les technologies de communications électroniques, le tourisme et le développement d'entreprises industrielles et tertiaires. Les équipes du GREYC seront impliquées dans chacune des parties du projet, en l'occurrence la numérisation intelligente, l'imagerie 3D et réalité virtuelle et la protection des contenus, sécurisation des échanges et technologies sans contact.

D'un point de vue international, les efforts importants réalisés pour publier dans des revues et colloques d'excellence se poursuivront sur le prochain quadriennal. Ils seront accompagnés d'une politique plus ambitieuse de collaboration scientifique avec les meilleures équipes de nos domaines de recherche. Le GREYC n'est pas très actif dans le montage Européen et intervient en tant que participant de manière marginale. Compte tenu de la masse de travail que cela représente, il n'est pas opportun que notre laboratoire coordonne le montage et la gestion de tels projets. En revanche, la participation à des projets sera encouragée.

Les équipes du GREYC, conscientes de l'évolution du paysage de la recherche local, national et international, ont préparé les projets d'équipe décrits dans les chapitres suivants. L'organisation et la gouvernance du laboratoire doivent également s'adapter à ces changements. Ces évolutions, décrites dans les sections suivantes, sont marquées principalement par la reconfiguration de ses équipes pour une meilleure lisibilité scientifique, la création de départements scientifiques, la création de plateformes et la création d'un conseil scientifique. La description des plateformes fait l'objet d'un chapitre spécifique de ce document.

Structuration du laboratoire

Durant le prochain quadriennal, le GREYC sera composé d'*équipes* de recherche organisées en *départements*, de services communs, de plateformes et disposera pour sa gouvernance d'un *conseil de laboratoire*,

3. Maison de la Recherche en Sciences Humaines de l'Université de Caen

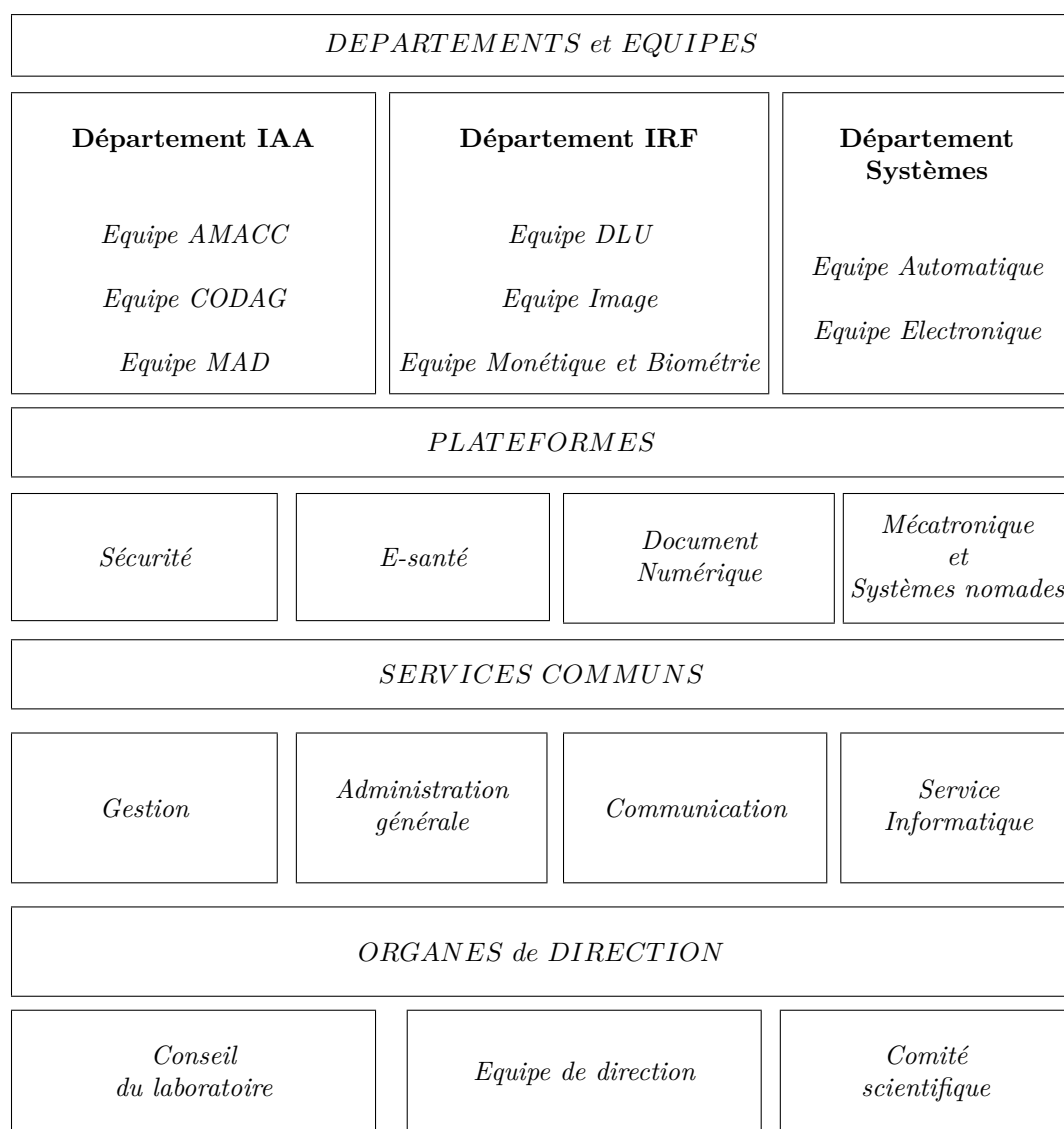


FIGURE 1 – Organisation du laboratoire

d'un comité scientifique et d'une équipe de direction. Le lien entre ces éléments est montré par la figure 1.

Les équipes de recherche. Le GREYC est attaché à ce que la recherche se fasse au sein d'équipes de recherche. Une équipe est un groupe de personnes clairement identifiées (typiquement de 5 à 15 permanents) qui poursuivent des objectifs de recherche communs et partagent la même vision pour y parvenir. Chaque chercheur ou enseignant-chercheur du laboratoire est rattaché à une équipe. Une équipe est placée sous la responsabilité d'un *responsable d'équipe* élu. Ce responsable assure une animation scientifique et veille à ce que chacun trouve sa place dans l'équipe. Le responsable d'équipe doit également avoir une très bonne visibilité de l'ensemble de la recherche conduite dans le laboratoire afin d'aider l'équipe à bien se positionner au sein du laboratoire. Chaque équipe se structure en *thèmes* de recherche, lesquels constituent un projet de recherche cohérent accompagné d'une stratégie pragmatique pour sa réalisation.

Comme il a été annoncé dans le bilan, le GREYC sera composé de 8 équipes durant le prochain quadriennal.

- L'équipe Algorithmique, Modèles de Calcul, Aléa, Cryptographie, Complexité (AMACC)
- L'équipe *Contraintes, Data Mining, Graphes (CODAG)*
- L'équipe *Modèles, Agents et Décision (MAD)*
- L'équipe *Document numérique, Langue, Usages (DLU)*
- L'équipe *Image*
- L'équipe *Monétique et Biométrie*
- L'équipe *Automatique*
- L'équipe *Electronique*

Le renouvellement des équipes. Une unité de recherche comme la nôtre doit pouvoir s'adapter aux évolutions qu'il s'agisse des évolutions des personnels ou des thèmes de recherche. Il peut donc être opportun qu'à certains moments des équipes disparaissent, apparaissent ou se fusionnent. A l'avenir, les porteurs d'une nouvelle équipe, qu'il s'agisse d'une équipe créée ex nihilo ou par fusion d'équipes existantes devront rédiger un projet de l'équipe : un document présentant outre les objectifs scientifiques, un positionnement local, national et international des travaux, ainsi que les moyens mis en œuvre pour atteindre les objectifs scientifiques visés. Ce projet sera présenté, discuté et affiné lors d'une réunion du comité scientifique du laboratoire, puis le conseil du laboratoire donnera son avis sur l'opportunité de créer l'équipe. C'est sur cet avis motivé que la direction du laboratoire prendra la décision de créer l'équipe.

Les services communs. Les services communs constituent un soutien à la recherche sur ses différentes facettes : technique, administrative, gestion et communication.

- Le *service informatique* assure la gestion, l'installation et la maintenance du parc informatique du laboratoire.
- Le *service administration générale* assure la gestion administrative du laboratoire et le lien avec les tutelles.
- Le *service gestion* assure la gestion du budget en général en terme de projets, de contrats, de commande et des recrutements de personnels contractuels.
- Le *service communication* assure la mise à jour du site du laboratoire, le lien avec la presse et les médias et la reprographie.

Les plateformes. Les plateformes sont de nouveaux instruments pour la réalisation de projets scientifiques pluridisciplinaires. Les plateformes regroupent sur un site unique des ressources scientifiques et/ou technologiques de haut niveau et des moyens humains, e.g. ingénieurs, accessibles à une communauté d'utilisateurs. Elles peuvent favoriser l'émergence de projets communs à plusieurs équipes ainsi que la valorisation de leur travaux de recherches et l'initiation de collaborations avec des entreprises ou autres structures, notamment le CEA et la DGA. Les quatre plateformes suivantes seront réalisées :

- La plateforme sécurité.
- La plateforme document numérique.
- La plateforme e-santé.

– La plateforme mécatronique et systèmes nomades.
Elles seront décrites dans le chapitre suivant.

Départements scientifiques. Compte tenu du nombre important d'équipes de recherche et de leur diversité thématique, une organisation en 3 départements scientifiques sera adoptée. Elle permettra une meilleure lisibilité des activités du laboratoire. En plus de la lisibilité, nous recherchons au travers de ces départements à asseoir une culture scientifique commune, en favorisant les échanges entre les équipes et l'émergence de projets innovants.

Les 3 départements seront :

- Le département *Intelligence Artificielle et Algorithmique (IAA)* comprenant les équipes AMACC, CODAG et MAD. Dans ce département, les équipes partagent une culture scientifique commune en termes d'outils formels (logique, complexité, graphes, probabilités, ...), d'outils de modélisation (problèmes de satisfaction de contraintes, théorie de la décision, modélisation de l'aléa, automates cellulaires, systèmes dynamiques, systèmes multi-agents, ...) et de méthodes algorithmiques (algorithmique de l'approximation, algorithmique probabiliste, optimisation déterministe ou stochastique, programmation linéaire, programmation dynamique, ...)
- Le département *Information et Reconnaissance des Formes (IRF)* autour des équipes Image, Monétique et Biométrie et DLU. Le point commun de ces équipes est celui de l'utilisation et du développement de modèles pour le traitement de l'information, ce qui constituera le cœur de ce département.
- Le département *Systèmes* comprenant les équipes Automatique et Electronique. Le fil conducteur de ce département est l'approche système qui y est développée. Nous espérons que l'existence de ce département contribuera à faire émerger des projets de recherche communs sur une instrumentation avancée, exploitant les compétences de l'équipe Electronique en conception des capteurs physiques et celles de l'équipe Automatique en synthèse des capteurs logiciels et en modélisation des signaux et systèmes.

Les activités d'un département, e.g. séminaires du département et projets communs, sont complètement définies par les responsables des équipes qui le constituent. Un animateur du département est désigné par les responsables des équipes pour veiller à la réalisation de ces activités.

Comité scientifique. Le comité scientifique sera par excellence le lieu d'échanges scientifiques du laboratoire comme le fut le comité de direction actuel. Il renforcera son identité et élaborera ses orientations et sa stratégie de recherche. Composé au minimum du directeur et des directeurs adjoints ainsi que des huit responsables des équipes, il sera complété par d'autres membres du laboratoire (typiquement 5) nommés par la direction en concertation avec les responsables des équipes, et par des membres extérieurs au laboratoire (typiquement 2). Le comité scientifique sera présidé par l'un des directeurs adjoints qui en assurera la coordination et l'organisation. Le comité scientifique se réunira mensuellement pour travailler sur des missions d'étude concernant des sujets scientifiques définis à l'avance sur la base d'un calendrier précis. Il aura également la possibilité de nommer des commissions d'évaluation de projets de recherche du laboratoire, de donner un avis sur les équipes et de réaliser une auto-évaluation continue du laboratoire. Le comité scientifique permettra de renforcer l'identité du laboratoire, grâce à une meilleure connaissance mutuelle des équipes et de leurs travaux.

Conseil du laboratoire. Le conseil du laboratoire est l'organe statutaire du laboratoire. Il sera consulté par la direction du laboratoire sur la coordination des recherches, la composition des équipes, les moyens budgétaires et leur répartition, la politique des contrats de recherche et de la valorisation, la politique de diffusion de l'information scientifique et de la communication, la gestion des ressources humaines, la politique de formation par la recherche, et plus généralement sur toutes les mesures liées à l'organisation et au fonctionnement du laboratoire. Le conseil de laboratoire sera en outre tenu informé par le directeur du laboratoire de la politique des tutelles. Composé de 30 membres, il sera présidé par le directeur du laboratoire. Il comportera des membres élus, des membres de droit (responsables des équipes, directeurs adjoints) et des membres nommés pour assurer une répartition équilibrée. Le conseil du laboratoire aura une implication plus forte dans la prise des décisions lors du prochain quadriennal. Il se réunira une fois par trimestre pour étudier les diverses questions relatives à la politique scientifique,

la gestion des ressources, l'organisation et le fonctionnement du laboratoire. Tous les avis seront soumis à des votes.

L'équipe de direction. Le GREYC est une grosse unité qui doit prendre des décisions nombreuses et parfois complexes. Il paraît opportun que le directeur s'appuie sur 2 directeurs adjoints lors du prochain quadriennal (contre 1 directeur adjoint jusqu'à présent). Tout en assurant plus facilement une présence continue de la direction dans le laboratoire pour la gestion quotidienne des deux sites, une direction tripartite permettra plus de concertation lorsqu'il s'agit de prendre des décisions rapides. Une direction à 3 permettra également une meilleure répartition de tâches prenantes telles que la représentation du laboratoire auprès des instances régionales ou nationales, l'animation du comité scientifique, etc.

L'équipe de direction s'efforcera de

- Conforter la visibilité de l'unité, que ce soit vis à vis de nos établissements de tutelle, de la région et des pôles de compétitivité, et plus généralement vis à vis des dispositifs de recherche nationaux ou internationaux. Par exemple, la direction établira un agenda annuel des rencontres avec ses tutelles.
- Définir une feuille de route avec les équipes avec un agenda annuel de rencontres qui permettront de faire régulièrement les bilans scientifiques de chaque équipe, et, sur la base de ses résultats et de son rayonnement scientifique, d'évaluer ses besoins et de l'aider à redéfinir ses orientations si besoin est.
- Veiller à la qualité des services communs et au suivi des carrières des ITA et IATOS.
- Conforter la culture de laboratoire en veillant à ce que toutes les sensibilités soient entendues, que chacun puisse s'exprimer et soit heureux et fier d'être membre du laboratoire.

Les décisions prises par la direction seront élaborées à partir des travaux conduits dans le comité scientifique et des avis du conseil du laboratoire. Concernant la politique de répartition des moyens, l'unité de recherche continuera à adopter une conception solidaire qui privilégie la mutualisation des ressources et la prédiction des besoins.

Membres producteurs / non producteurs. Pour préparer le prochain quadriennal, les responsables d'équipes ont procédé à un recensement de leurs membres dits *non producteurs* ou *insuffisamment producteurs*. Ils les ont rencontrés individuellement pour cerner leurs difficultés et les motiver à recouvrer une dynamique de recherche fructueuse. La direction du laboratoire avec les responsables d'équipes a invité les membres non producteurs qui se sont trop investis dans des activités pédagogiques et des responsabilités collectives à rééquilibrer leurs activités dans un sens plus favorable à la recherche. Des résultats prometteurs de cette politique incitative ont d'ores et déjà été obtenus. Pour le prochain quadriennal, il a été demandé à tous les membres du laboratoire d'avoir une attache explicite avec le laboratoire. Cette attache est définie par une volonté d'adhérer au laboratoire et se traduit par des productions scientifiques. L'appartenance au GREYC est exclusive et n'est acquise qu'en échange d'un engagement à contribuer à sa vie scientifique, en y conduisant un projet scientifique cohérent avec les thèmes de recherche de l'unité. L'appartenance est validée par le responsable d'équipe et par la direction de l'unité.

Nous distinguons les membres *producteurs* des membres *non-producteurs* sur la base des critères définis par l'AERES modulés par des critères d'évaluation propres aux équipes. Les membres considérés comme *non-producteur* ont reconnu cet état de fait et ont marqué leur volonté d'appartenance à l'unité en définissant un projet de recherche avec leur responsable d'équipe et en s'engageant à le réaliser au cours du prochain quadriennal, comme cela est explicitement indiqué dans leurs fiches individuelles. Le projet de recherche est conçu à partir des connaissances fondamentales du membre concerné et repose sur sa capacité à s'impliquer d'une manière effective dans au moins un thème de l'équipe. Il a été demandé à tous les membres non-producteur souhaitant un rattachement au laboratoire de :

- Réduire leurs responsabilités administratives, si celles-ci font obstacle à la réalisation de leur projet de recherche.
- D'actualiser leurs connaissances fondamentales via un état de l'art sur leur projet de recherche, étape importante pour recouvrer efficacement une dynamique de recherche.
- De s'impliquer de manière effective dans la vie scientifique du laboratoire par des présentations aux séminaires d'équipe ou de département, ainsi qu'aux projets collaboratifs de l'équipe (ANR, contrat de collaboration, etc.).

- De participer activement à des encadrements des stagiaires, aux comités d'encadrement des thèses et aux projets collaboratifs de l'équipe (ANR, contrats de collaboration, etc.).
- De s'impliquer dans des GDR et d'y confronter périodiquement leurs idées avec la communauté scientifique.

De son côté, la direction s'engage à accompagner les non producteurs vers une bonne dynamique de recherche en leur affectant des ressources si nécessaire.

Vie du laboratoire

Les journées du GREYC. Nous organiserons plusieurs journées pour accueillir les nouveaux arrivants et permettre aux membres de l'unité de se retrouver et échanger autour de leurs activités de recherche .

Notons en particulier :

- La *journée d'accueil des nouveaux* pour leur permettre de faire connaissance avec l'unité et son personnel. Cette journée a été d'ores et déjà mise en place depuis 2009.
- La *journée des doctorants à mi-parcours*, journée durant laquelle les doctorants à mi-parcours présenteront leurs travaux. Tout en permettant de contrôler le bon déroulement des thèses, cette journée contribuera à enrichir la connaissance mutuelle que nous avons de nos travaux et ainsi renforcer l'identité de l'unité.
- La *journée consacrée à la présentation de résultats majeurs* présentés par des membres du laboratoire et en la présence d'un scientifique de renom.
- Des *journées spécifiques* destinées à nos tutelles, aux pôles de compétitivité, à la région et aux industriels, organisées dans le but de leur présenter nos travaux de recherche.

La journée des doctorants et celle de présentation des résultats majeurs seront organisées à l'extérieur du laboratoire, dans une ambiance conviviale et propice à favoriser les échanges.

Durée des thèses. Le laboratoire est particulièrement soucieux de l'avenir de ses doctorants et du respect de la charte de thèse qui prévoit une durée qui ne dépasse pas 40 mois. La durée moyenne des thèses au GREYC est sensiblement supérieure à 40 mois, aussi nous conforterons les dispositions qui ont été d'ores et déjà mises en place en vue de réduire cette durée :

- La création d'un comité de suivi de thèse pour faire le point sur l'état d'avancement des thèses à l'issue de la journée des doctorants à mi-parcours. Ce comité fera des recommandations aux directeurs de thèses si besoin est.
- Les directeurs des thèses dont la durée des thèses sera anormalement élevée devront justifier cette durée.
- Concernant l'affectation des postes d'ATER, la priorité sera accordée aux candidats ayant soutenu ou ayant terminé la rédaction de leur thèse. Cette disposition qui a été d'ores et déjà mise en place depuis 2009 a permis d'éviter de charger le doctorant au moment où il doit se consacrer à la rédaction de sa thèse. Les contrats doctoraux d'enseignement ne seront donnés pour cette même raison qu'aux doctorants de 1ère et 2ème année.

Actions incitatives. La direction développera les actions incitatives qui ont été d'ores et déjà entreprises pour bien asseoir une culture de l'excellence au sein du laboratoire :

- Proposer des candidats aux distinctions nationales (prix de thèse et médailles, etc.),
- Mettre en place un budget spécifique pour l'aide au montage des projets internationaux et européens et la présence dans les réseaux d'excellence, en finançant si cela s'avère nécessaire le recours à des cabinets privés spécialisés pour ce type de montage.
- Inciter à passer une HDR, pour ceux dont le dossier le permet mais qui tardent à le faire. Des résultats probants ont été obtenus lors du dernier quadriennal (7 HDR soutenues entre 2006 et 2010 et 8 dossiers de candidature HDR déposés en 2010).
- Inciter aux délégations et CRCT pour la recherche. Cette action a été fructueuse sur le dernier quadriennal puisque le GREYC a bénéficié de 8 délégations (4 CNRS, 2 INRIA et 2 IRD) et 6 CRCT (3 par la voie locale et 3 par la voie nationale). Ces délégations permettent de consolider des axes de recherche du laboratoire et de redéfinir certaines thématiques si besoin est.
- Favoriser les manifestations mettant en avant les résultats phares du laboratoire.

- Organiser un séminaire (ouvert sur l'extérieur) présentant des contributions et des faits marquants du laboratoire en plus des séminaires réguliers des équipes et des départements.
- Soutenir toutes les activités des membres en matière de diffusion et de vulgarisation, notamment la participation aux journées de la Fête de la Science et aux manifestations du type *portes ouvertes*.

Politique de recrutement et d'emploi. Nous comptons poursuivre la politique de recrutement introduite dans le précédent quadriennal, politique qui vise à recruter de manière préférentielle des personnes extérieures au laboratoire. Cela nous paraît être un élément indispensable au renouvellement des idées et au dynamisme d'un laboratoire. En 2010, les trois enseignants-chercheurs recrutés sont des extérieurs qui possèdent d'excellents dossiers, tant d'un point de vue de la recherche que de celui de l'enseignement. La future direction du GREYC est consciente de l'importance des recrutements pour la dynamique du laboratoire et ses ambitions d'excellence et de rayonnement. Elle veillera à développer une communication efficace afin d'attirer les meilleures candidatures. Elle compte également sur le soutien de nos tutelles en terme d'emplois, particulièrement de rang A, afin de pouvoir assurer un meilleur encadrement scientifique des jeunes chercheurs et des doctorants.

Huit départs à la retraite d'enseignants-chercheurs sont prévus dans le prochain quadriennal. Nous engagerons une réflexion en profondeur afin d'élaborer un plan de recrutement cohérent avec le projet scientifique du laboratoire. La politique sera de maintenir ces postes au sein du laboratoire, parfois en les redéployant au travers des équipes, comme ceci a été réalisé au précédent quadriennal.

En ce qui concerne les chercheurs, nous chercherons à attirer des candidats extérieurs, comme nous avons su le faire dans le passé.

En ce qui concerne les personnels ITA-IATOS, la situation est plutôt inquiétante en dépit des trois postes CNRS en redéploiement (NOEMI) et du poste de technicien créé à la demande de l'ENSICAEN. Les services communs du laboratoire seront réorganisés pour une répartition des tâches plus cohérente avec la nouvelle structure du laboratoire. Les fiches de postes seront clairement définies à partir d'une mutualisation optimale des ressources humaines sur les deux sites.

Notre unité ne pourra avoir des services communs fonctionnels que si des postes sont ouverts sur les priorités suivantes :

- Un poste de secrétariat supplémentaire est crucial, afin d'assurer aux côtés de l'assistante de direction une gestion centralisée des budgets et de la communication du laboratoire.
- Deux postes d'IE pour les équipes Automatique et Electronique sont nécessaires pour conforter leurs activités de recherche méthodologique et de valorisation.
- Deux postes d'IE en Informatique sont nécessaires au développement et à la maintenance des plates-formes et la conservation du patrimoine logiciel du laboratoire.

Auto-analyse

Conforter nos points forts

- Une bonne production scientifique, tant par sa quantité que par sa qualité, tant sur les aspects fondamentaux que sur les aspects méthodologiques et appliqués.
- Une bonne insertion locale avec une forte implication dans les pôles de compétitivité et les structures locales.
- Une très bonne activité contractuelle et de valorisation industrielle (contrats de collaboration scientifique, contrats CIFRE, Start-up incubées, etc). Durant le dernier quadriennal, l'unité a été impliquée dans 28 projets ANR, dont 16 projets en collaboration avec des industriels, et 21 contrats de collaboration scientifique avec des industriels.
- Une recherche fondamentale partenariale reconnue comme en témoignent les 6 projets ANR blancs labellisés entre 2006 et 2009.
- Une bonne ambiance de travail.
- Des services communs de qualité malgré le manque notable de personnel IATOS.
- Une forte implication du laboratoire dans les activités de formation avec des retombées positives sur la recherche.
- Le développement de notre rayonnement scientifique par l'organisation de conférences et d'écoles thématiques internationales.

Surmonter nos points faibles

- Tirer profit de la nouvelle structuration des équipes et de l'organisation en départements pour favoriser l'émergence de projets fédérateurs innovants.
- Réduire le nombre de non-productifs. Des résultats prometteurs ont été obtenus sur le dernier quadriennal et nous espérons faire mieux sur le prochain, par la mise en place de la politique décrite ci-dessus.
- Accroître le rayonnement national et international de l'unité par une politique de publication et de collaboration scientifique plus agressive.
- Réduire la durée de thèses via les comités d'encadrement et de suivi de thèses.
- Conforter les efforts de communication pour une meilleure visibilité du laboratoire.
- Renforcer la politique incitative des membres du laboratoire à soutenir une HDR afin d'accroître les capacités d'encadrement de l'unité.
- Accroître nos collaborations industrielles ; nous pourrions certainement profiter davantage de notre succès en matière de projets ANR pour augmenter notre visibilité auprès des industriels.
- Développer nos collaborations européennes et utiliser nos réseaux d'excellence pour participer davantage à de nouveaux contrats européens.

Opportunités

- L'organisation en départements scientifiques et la nouvelle structuration des équipes devraient permettre de renforcer la cohérence scientifique du laboratoire.
- La nouvelle équipe DLU a toutes les compétences requises pour permettre au laboratoire d'être le leader du projet document numérique de la région, et un acteur incontournable de la scène nationale et internationale. Le projet Numérique Normandie déposé dans le cadre du Grand Emprunt est une excellente opportunité pour développer le thème Document Numérique.
- L'intérêt des pôles TES et MOVE'O aux systèmes nomades et à la mécatronique est susceptible d'amplifier les efforts de recherche que nous faisons dans ces domaines.
- Les projets labellisés par les pôles de compétitivité pourront être utilisés pour intensifier notre valorisation industrielle.
- Nos collaborations internationales, notamment avec les pays européens, et notre participation aux réseaux d'excellence permettront d'accroître notre attractivité et favoriseront l'émergence de projets européens.
- Les plates-formes permettront une meilleure visibilité des réalisations issues de nos recherches, tout en structurant et développant notre partenariat avec la région.
- Les besoins en enseignement d'informatique pourraient favoriser des recrutements sur les thématiques fondamentales du laboratoire et l'émergence de nouvelles thématiques.
- Nos projets ANR blancs confortent le développement du vivier des doctorants motivés par une recherche fondamentale.

Risques

- Privilégier des projets de recherche fédérateurs collaboratifs peut nuire aux projets de recherche structurants des équipes et faire perdre leur identité aux équipes.
- La politique actuelle en matière de création et redéploiement de postes, relativement ouverte sur l'extérieur, pourrait induire une démobilité des enseignants-chercheurs aspirant à des promotions internes.
- Le potentiel d'encadrement scientifique est faible dans certaines équipes, des recrutements en collègue A seront nécessaires pour conforter certains thèmes.
- La période de transition avant le début du prochain quadriennal est relativement difficile pour l'équipe DLU à cause des départs prévus.
- Une politique de valorisation industrielle plus agressive pourrait nuire à certaines équipes dont l'activité est centrée sur une recherche très fondamentale.

Plateformes

Porté par l'ambition de maintenir et développer son activité de transfert technologique et de valorisation, le GREYC a décidé de se doter de quatre plateformes. Comme nous l'avons expliqué précédemment, il s'agit de regrouper en un site unique des ressources scientifiques et/ou technologiques de haut niveau et des moyens humains, e.g. ingénieurs, accessibles à une communauté d'utilisateurs. Outre les aspects de valorisation, ces plateformes visent aussi à favoriser l'émergence de projets transverses et l'initiation de collaborations avec des entreprises ou d'autres structures comme le CEA et la DGA.

Si les plateformes sont essentielles comme vitrines du savoir faire du laboratoire, elles permettent également de pérenniser les travaux réalisés par des personnels non permanents, e.g. doctorants, post-doctorants et ingénieurs projet.

Le développement et la maintenance de ces plateformes requièrent un effort conséquent en moyens humains (ingénieurs d'étude et ingénieurs de recherche en particulier) ; deux ingénieurs d'étude permanents sont demandés pour ces besoins.

0.1 Plateforme sécurité

Cette plateforme comprendra différentes composantes allant de la validation et du test des produits issus de recherches finalisées à l'accompagnement de la recherche fondamentale.

- La composante *monétique matérielle*, constituée de produits industriels, permettra le test de nouvelles solutions de paiement développées au sein du GREYC dans le cadre de projets collaboratifs industriels.
- La composante *biométrique* rassemblera le savoir-faire du GREYC dans ce domaine et permettra l'évaluation de systèmes biométriques, tant du point de vue de leur performance, que de leur sécurité et leur acceptabilité.
- Le logiciel *GREYC Keystroke* permettra la création de benchmarks pour l'authentification par dynamique de frappe au clavier.
- La composante *IOTA* réunira la gestion des événements liés aux déplacements de marchandises et des échanges de documents entre entreprises (B2B).
- La composante *cryptologie* développera des programmes pour la cryptanalyse et des implantations d'algorithmes arithmétiques ou issus des mathématiques discrètes pour la conception et l'évaluation de nouveaux cryptosystèmes.

Cette plateforme et ses multiples facettes sont un atout essentiel en relation avec le pôle de compétitivité Transactions Electroniques Sécurisées situé à Caen et les formations dispensées à l'UCBN (Master AMI - Parcours Sécurité, futur master eSecure) et à l'ENSICAEN (les différents cursus en Monétique et Sécurité Informatique). Par ailleurs, elle répond naturellement aux remarques de la DGRI⁴ dans son enquête sur la recherche académique française en sécurité informatique en novembre 2007, qui souligne l'intérêt de l'émergence de plateformes en sécurité informatique en France.

4. Direction Générale pour la Recherche et de l'Innovation

0.2 Plateforme e-santé

La création d'une plateforme e-santé au sein du GREYC a pour vocation de renforcer et de fédérer toutes les activités dans le domaine de la santé. L'ultime motivation est de proposer aux médecins, biologistes et pathologistes des outils d'aide à la décision efficaces dans le domaine de la santé. Une attention particulière sera accordée aux ouvertures technologiques innovantes propres à l'imagerie médicale numérique. Ce projet de recherche s'appuie sur les trois atouts suivants de notre unité.

- Une recherche multidisciplinaire et compétitive en imagerie médicale et biomédicale menée au sein du GREYC depuis plus de 15 ans.
- Une stratégie scientifique régionale soutenue autour des thématiques de la santé. Cette stratégie a été d'abord élaborée dans le cadre du pôle régional Traitement et Analyse d'Images et du pôle ITIC⁵ du CPER 2000-2006 et est actuellement confortée dans le cadre de l'IRCBN⁶ et du projet ARCHADE autour de l'Hadronthérapie.
- Une stratégie de valorisation et d'intérêt économique régional. Plusieurs résultats de recherche du GREYC et de ses partenaires médecins ou pathologistes ont fait l'objet de transferts et de collaborations industrielles avec des sociétés régionales ou nationales.

Les perspectives des travaux de recherche qui seront développés dans le cadre de la plateforme e-santé porteront essentiellement sur l'imagerie médicale et biomédicale et la fouille des données médicales. Dans ce dernier domaine, une attention particulière sera accordée au programme MultiPULS⁷ pour la veille épidémiologique multilingue.

Les projets de recherche sur l'imagerie médicale et biomédicale concernent les domaines de l'imagerie du cerveau et l'imagerie en biologie et en pathologie dans une optique d'aide au diagnostic du cancer. La fouille de données médicales s'inscrit dans le cadre d'une collaboration avec le LCMT⁸ sur l'étude des propriétés chimiques des molécules pour la santé, notamment les motifs fréquents dans des bases de molécules. Ces derniers permettent d'élaborer des modélisations 3D et des schémas de réaction des molécules. On peut ainsi évaluer l'impact sanitaire et environnemental des relargages au sein d'écosystèmes sensibles comme les estuaires et réaliser une estimation de la toxicité.

0.3 Plateforme mécatronique et systèmes nomades

La création d'une plateforme *mécatronique et systèmes nomades* vient en synergie avec celle de la filière mécatronique de l'école d'ingénieur ESIX de l'UCBN, projet auquel les équipes Automatique, Électronique et MAD du GREYC ont contribué activement. D'un point de vue local, cela reflète l'importance des systèmes nomades et de la mécatronique dans les nouvelles orientations des pôles de compétitivité TES et MOV'EO. Les bases fondamentales requises pour la réalisation d'une telle plateforme sont disponibles au sein des équipes de l'unité. Nous distinguons en particulier :

- Les compétences reconnues de l'équipe Électronique sur la mise au point et l'optimisation de dispositifs électro-mécaniques et le contrôle non destructif par courant de Foucault. Un capteur électro-mécanique performant du champ magnétique est en cours de réalisation dans le cadre du projet de recherche américain "Heterostructural Uncooled Magnetic Sensors" de la DARPA⁹ dont le GREYC est partenaire. Des méthodes appropriées pour l'analyse de la qualité des pièces métalliques et de soudure de manière non-invasive ont été développées, elles constituent un enjeu capital pour les industries automobile et aéronautique.
- Les compétences fondamentales reconnues de l'équipe Automatique en matière de modélisation des signaux, de conception d'asservissements robustes et de synthèse des observateurs pour les systèmes non linéaires. Ces compétences sont vitales pour la conception d'une instrumentation fiable à haute valeur ajoutée. Des capteurs logiciels ont d'ores et déjà été conçus pour réaliser une

5. Image et Technologies de l'Information et de la Communication

6. Institut Régional du Cancer de Basse Normandie

7. Multilanguage Pattern based Understanding and Learning Systems

8. Laboratoire de Chimie Moléculaire et Thio-organique

9. Defense Advanced Research Projects Agency

estimation consistante de la vitesse d'un moteur asynchrone à partir des mesures des courants statoriques. Une validation expérimentale réussie a été effectuée à partir d'une mise en œuvre sur une carte DSP. Des méthodes de filtrage appropriées ont été développées pour traiter le cas des bruits de mesure relativement importants par rapport au signal.

- Les compétences de l'équipe MAD sur la mise en œuvre des systèmes intelligents et autonomes. Ces compétences ont été utilisées avec succès dans plusieurs projets de recherche concernant les systèmes robotisés. Des perspectives d'interaction entre systèmes autonomes et personnes sont développés dans le cadre de projets de recherche avec des applications potentielles dans les domaines de la robotique d'assistance, du véhicule intelligent et des drones. Des démonstrateurs pour les robots autonomes, les systèmes multi-robots et l'interaction simple homme-robot sont disponibles.

La fédération de ces compétences nous permettra de développer toutes les composantes d'une instrumentation à hautes performances et de systèmes de commande avancée pour les systèmes électromécaniques, en l'occurrence les capteurs pour faire les mesures requises, les observateurs pour la conception de capteurs logiciels, une approche expérimentale pour l'estimation optimale, et les outils de programmation adéquats pour la mise en œuvre.

Il sera ainsi possible d'initier une synergie prometteuse entre les équipes Automatique, Électronique et MAD pour la conception de systèmes embarqués spécifiques à l'instrumentation et la commande de robots. La commande de robots permettrait de développer les démonstrateurs de l'équipe MAD dans le cadre d'une collaboration avec les équipes Automatique et Image pour la vision robotique.

0.4 Plateforme document numérique

Les travaux de recherche menés au GREYC autour du document numérique couvrent des aspects multiples, depuis l'aide à la conception de documents électroniques, jusqu'à la navigation dans de grandes collections comme les bibliothèques numériques et l'indexation de documents dans ces bibliothèques ou sur Internet, en passant par la fouille de données textuelles ainsi que le tatouage de documents et la sécurité des échanges. Outre une recherche fondamentale sur ces différents aspects, des plateformes logicielles ont été développées dans le cadre de projets menés en collaboration avec des chercheurs en sciences humaines et des partenaires industriels.

La finalité de la plateforme document numérique est de fédérer toutes les compétences reconnues du GREYC autour de la valorisation des documents dans les domaines de la linguistique, de la fouille de documents et de l'interprétation des images. Une attention particulière sera accordée aux thématiques de recherche suivantes.

- Le traitement automatique des textes avec des méthodes différentielles. Ces méthodes ont l'avantage d'être multilingues et sont applicables à tout type de document, y compris des pages web. L'actuelle plateforme WIMS pour le recueil, l'analyse et la diffusion de documents exploite différentes formes de documents et est indépendante des langues.
- La constitution de ressources pour le français et l'analyse sémantique avec ressources dans la perspective de construction du web sémantique. Les plateformes d'expérimentation *LinguaStream* en traitement automatique des langues et la plateforme *GLOZZ* pour l'annotation de corpus de discours seront utilisées dans ce but.
- L'interprétation de documents multimédia pour l'exploitation automatique du texte et de l'image des dossiers médicaux, des ouvrages pédagogiques, e.g. les atlas de géographie, et l'aide à la recherche d'information dans les livres anciens. Cet aspect concerne aussi la recherche sur Internet de documents composites et multimédia comme les vidéos. La transposition de documents écrits en documents audio ou haptiques s'inscrit dans le cadre de la lutte contre le handicap.
- La fouille de textes pour l'extraction de connaissances à partir d'une hybridation prometteuse des méthodes de traitement automatique des langues et de fouille de données. Les applications concernent la recherche fine d'information dans les bibliothèques numériques, notamment en médecine, ou le diagnostic et l'adaptation automatique des extracteurs de mots-clés pour la surveillance de flux de documents pour la veille épidémiologique par exemple. Une plateforme d'élaboration de scénarios de fouille de données s'appuyant sur une interface de programmation graphique est en

cours de développement.

Cette plateforme favorisera la synergie entre les équipes CODAG, DLU et Image. Elle permettra au GREYC de renforcer son rôle dans le projet *Normandie Numérique* porté par l'Université de Caen et de rayonner sur la région. On notera que la plateforme d'expérimentation LinguaStream a donné lieu à la création d'une start-up en 2008 et qu'un nouveau projet à partir de la plateforme Wims est en incubation.

1

Équipe Algorithmique, Modèles de calcul, Aléa, Cryptographie, Complexité

1.1	L'équipe	15
1.1.1	Les trois thèmes.	15
1.1.2	Les membres.	15
1.1.3	Le contexte de la création de l'équipe AMACC.	16
1.1.4	La politique scientifique de l'équipe AMACC.	17
1.1.5	Risques et opportunités.	18
1.2	Thème 1 : Modèles de calcul, logique et complexité.	18
1.3	Thème 2 : Protection de l'information, codage et cryptographie.	20
1.4	Thème 3 : Structures aléatoires et analyse d'algorithmes.	22
1.5	Les projets de l'équipe en cours en 2011.	24

Equipe AMACC

Algorithmique, Modèles de calcul, Aléa, Cryptographie, Complexité

Responsable : Brigitte VALLÉE

L'équipe Algorithmique se reconfigure après le départ de son Thème 4. Elle donne naissance à la nouvelle équipe AMACC. Cette équipe se retrouve d'abord autour de deux concepts génériques, l'algorithme et la complexité, symbolisés par le premier A et le dernier C, qui sous-tendent toutes ses activités. Elle y adopte des points de vue complémentaires. Le premier étudie les modèles de calcul ou la notion de complexité, dans le pire des cas via les classes de complexité (le M pour Modèles de Calcul), tandis que le second travaille dans un cadre aléatoire, avec des modèles probabilistes (le second A pour Aléa). L'équipe a aussi une spécialité algorithmique bien marquée, et travaille en Protection de l'Information, notamment en codage et cryptographie (le premier C) : elle élabore des protocoles, analyse leur sécurité, ou les attaque, en s'appuyant sur ses connaissances fines en théorie de l'information et en arithmétique. Elle s'ouvre aussi à d'autres thématiques de la sécurité, comme l'internet des objets ou l'environnement embarqué sécurisé.

1.1 L'équipe

1.1.1 Les trois thèmes.

L'équipe s'organise selon trois principaux thèmes.

[T1] Modèles de calcul, logique et complexité

Responsable : Etienne GRANDJEAN

Etude des modèles de calcul, automates cellulaires et machines RAM, classes de complexité pour ces modèles, complexité des problèmes combinatoires ou logiques. Etude du paradigme de la programmation fonctionnelle pure.

[T2] Protection de l'information, codage et cryptographie

Responsable : Fabien LAGUILLAUMIE

Conception et analyse de la sécurité (via des preuves ou des attaques) de protocoles destinés à protéger l'information (cryptographie, codes correcteurs d'erreurs). Etude des objets et des algorithmes sous-jacents (issus de l'algèbre discrète ou de la théorie des nombres). Protection de la propriété intellectuelle, internet des objets, sécurité des réseaux.

[T3] Structures aléatoires et analyse d'algorithmes

Responsable : Brigitte VALLÉE

Analyse probabiliste des structures aléatoires et des algorithmes, avec un volet méthodologique notamment fondé sur la combinatoire analytique et les systèmes dynamiques, et deux volets plus applicatifs, où l'on étudie les mots et les nombres.

1.1.2 Les membres.

L'équipe comprend 13.5 membres permanents (un de ses membres se partage avec l'équipe DNLU « Document Numérique, Langue et Usages »), et 11.5 membres « producteurs ».

- 2 membres du collège A (1 DR CNRS, 1 PR à l'UFR Sciences),

- 11.5 membres du collège B (1 CR CNRS, 8.5 MC à l'UFR Sciences, 2 MC à l'ENSICAEN).

Elle comprend pour l'instant 3 doctorants, et va en recruter deux autres en Octobre 2010.

Prénom et nom	Fonction	T1	T2	T3	Date d'entrée
Ali Akhavi	MC UFR Sciences		X(-)	X(+)	2000
Julien Clément	CR CNRS			X	2005
Etienne Grandjean	PR UFR Sciences	X			1985
Jerzy Karczmarczuk	MC UFR Sciences	X			1988
Fabien Laguillaumie	MC UFR Sciences		X		2008
Jean-Marie Le Bars	MC UFR Sciences		X	X	1999
Loick Lhote	MC ENSICAEN		X(+)	X(-)	2007
Ayoub Otmani	MC ENSICAEN		X		2004 (**)
Jacques Madelaine (*)	MC UFR Sciences		X		1985
S. Ranaivoson	MC UFR Sciences	X			1991
Gaétan Richard	MC UFR Sciences	X			2009
Jean Saquet	MC UFR Sciences		X		1985
Véronique Terrier	MC UFR Sciences	X			1991
Brigitte Vallée	DR CNRS		X(-)	X(+)	1990
Johann Brault-Baron	A. Minist. + Mon	X			depuis 10/09
Mariya Georgieva	A. Minist.		X	X	depuis 10/09
Mathieu Roux	A. Minist. + Mon			X	depuis 10/07
Thu Hien Nguyen Thi	A. CNRS-Reg.			X	à partir de 10/10
Julien Devigne	A. CIFRE Orange Labs		X		à partir de 10/10

(*) Se partage avec l'équipe DLU « Document Numérique, Langue, Usages »

X(+) : rattachement principal – X(-) : rattachement secondaire

Elle accueille aussi deux autres membres non permanents :

– Sébastien Canard, ingénieur de recherche à Orange Labs (Caen), titulaire de l'habilitation à diriger des recherches, qui co-dirige la thèse de Julien Devigne.

– Marc Girault, cryptologue de stature internationale, ancien expert senior en cryptographie à FTR&D (Caen), titulaire de l'habilitation à diriger des recherches, maintenant professeur au Lycée Victor Hugo (Caen), qui participe aux enseignements de cryptologie du Master AMI, et apporte son soutien au thème T2 dans ses activités de veille scientifique.

1.1.3 Le contexte de la création de l'équipe AMACC.

Les quatre thèmes de l'ancienne équipe Algorithmique connaissent une évolution variable, plus ou moins importante, mais toujours notable, qui crée un nouveau contexte pour l'équipe AMACC.

Le départ du thème T4. Le thème T4 a quitté l'équipe, pour rejoindre le thème Fouille de Données de l'ancienne équipe Dodola et fonder la nouvelle équipe CODAG « Contraintes, DataMining, Graphes ». C'est une reconfiguration qui s'est préparée tout au long de ce quadriennal, par un resserrement continu des liens entre ces deux thèmes. Ce départ se fait dans de bonnes conditions humaines, et AMACC souhaite bonne chance à CODAG, et continuera à collaborer avec elle.

L'évolution sensible des thèmes T1 et T3. Les deux thèmes impairs évoluent dans le sens d'une meilleure insertion dans la nouvelle équipe. Le thème T1 cherche à faire converger les recherches sur les deux modèles de conception orthogonale (séquentiel et parallèle) qui sont étudiés en son sein. Le thème T3 réoriente les recherches qu'il menait sur le versant arithmétique en les élargissant à l'analyse des objets cryptographiques, et se rapproche encore davantage du thème T2.

La reconfiguration de T2. Même si son nom PICC est inchangé, le thème s'ouvre largement à d'autres thématiques que le codage et la cryptographie, et son statut à l'intérieur de AMACC est bien particulier. Tout ceci est dû à l'histoire récente et complexe de l'ancienne équipe SISTEM qui est décrite dans le chapitre « Sécurité » du présent rapport.

A l'occasion du projet pour le nouveau quadriennal, les membres du groupe PICC souhaitent se regrouper au sein du Thème 2 de l'Equipe AMACC, en y accueillant aussi d'autres activités reliées, comme la sécurité des objets ou des réseaux, ou la sécurité du document qui était retournée lors de la scission dans son équipe-mère, l'équipe Dodola, et qui y était assez isolée. Cela se traduit par des conséquences importantes :

(1) Le Thème 2 de l'équipe AMACC s'élargit par rapport à l'ancien Thème 2 de l'ancienne équipe Algo : il réunit les activités de protection de l'information du laboratoire et intègre, en particulier, à côté

du codage et de la cryptographie, des activités en sécurité liées à l'*internet des objets* et à l'embarqué. Cette réunion est une retombée positive de l'histoire de la sécurité au GREYC (voir chapitre « Sécurité »).

(2) Pour toutes les activités de la vie courante scientifique, le groupe PICC est abrité par l'équipe AMACC, et les trois jeunes maîtres de conférences qui en sont membres sont ainsi déchargés d'un certain nombre de tâches administratives, ce qui est appréciable pour eux. Le groupe PICC joue, au sein de AMACC, le rôle habituel d'un thème et peut ainsi continuer d'interagir de manière forte avec le Thème 3 de cette équipe.

(3) Afin de participer explicitement à la politique scientifique du laboratoire en matière de sécurité, le groupe PICC est représenté en tant que tel au comité scientifique du laboratoire.

(4) Quand le groupe PICC disposera de forces suffisantes, l'équipe AMACC sera heureuse de le voir prendre son indépendance, et quitter l'équipe. Cela ne pourra se produire que lorsque le laboratoire recrutera un professeur en protection de l'information.

On peut juger cette organisation un peu hybride. C'est pourtant, dans un environnement un peu difficile (voir chapitre « Sécurité »), la seule solution *transitoire* qui peut préserver les activités scientifiques du laboratoire en protection de l'information : c'est un *regroupement* des forces dans un même lieu, propice aux *interactions*, qui prend en compte son *sous-encadrement* : on le *décharge* des soucis de la vie quotidienne pour qu'il puisse mieux jouer son *rôle dans la politique scientifique* du laboratoire.

1.1.4 La politique scientifique de l'équipe AMACC.

La nouvelle équipe AMACC, issue du « rétrécissement » de l'ancienne équipe Algorithmique, a donc des effectifs plus faibles et des thématiques plus resserrées. Elle souhaite profiter de ces deux caractéristiques.

Renforcement du graphe des interactions internes. L'équipe vise d'abord à renforcer le graphe des interactions entre ses trois thèmes : L'arête T2–T3 est forte, mais les deux autres liaisons du triangle sont plus molles.

Arête T2–T3. Le bilan a prouvé la force des relations entre les thèmes T2 et T3, qu'il s'agisse des relations scientifiques ou plus largement des relations humaines de la vie de tous les jours. Elles seront encore renforcées, via l'évolution d'une partie des thèmes de T3, qui se rapproche de T2. Les collaborations sont liées, notamment, à l'existence de deux projets ANR *Lareda* et *Boole*, ou à la codirection de thèses, où les deux thèmes interviennent conjointement.

Arête T1–T3. Les interactions entre T1 et T3 se limitent pour l'instant au partage de la certitude culturelle de travailler sur les mêmes concepts : l'algorithme, la complexité. Jusqu'à présent, cette forte certitude ne s'est jamais vraiment concrétisée dans les faits, car les méthodes sont trop distantes. Le thème T1 utilise des outils de mathématiques discrètes (la logique, les graphes) et se concentre sur la complexité dans le pire des cas, tandis que le thème T3 utilise aussi des outils de mathématiques continues, travaille en théorie de l'information ou en arithmétique, et se concentre sur la complexité en moyenne. L'étude des classes de complexité probabilistes pourrait être un point de départ au rapprochement de ces thèmes.

Arête T1–T2. Il y a un membre de T1 vraiment intéressé par la cryptographie et très assidu au séminaire. Il pourrait faire le lien entre les deux thèmes...

Interactions entre AMACC et CODAG. Il n'y a aucune raison que les interactions qui existent actuellement

- entre les thèmes T1 et T4 de l'ancienne équipe Algorithmique, autour des notions de graphe et d'hypergraphe, de problèmes dans les bases de données,
 - entre les thèmes T3 et Fouille de Données des anciennes équipes Algorithmique et Dodola, sur l'étude de la complexité moyenne d'algorithmes de fouille de données,
- soient affectées par la reconfiguration des deux équipes. Ces collaborations, riches, seront poursuivies et renforcées.

Avenir du séminaire Algorithmique. La réussite du séminaire d'Algorithmique est due en partie à sa forte identité culturelle. Dans le futur, le séminaire conservera son nom. Tout en se centrant sur les thèmes scientifiques de l'équipe AMACC, qui en garde la responsabilité, il souhaite aussi couvrir, une partie du temps, les thématiques de l'équipe CODAG.

Rôle de l'équipe AMACC au sein du GREYC. Dans un laboratoire généraliste comme le GREYC, l'équipe AMACC peut jouer un rôle dans la sensibilisation des autres équipes aux problèmes de complexité. Elle

le joue un peu déjà pour l'analyse de la complexité dans le pire des cas, mais elle devrait faire de la publicité sur l'importance de l'analyse en moyenne. C'est plus difficile, car le type de mathématiques qui y est utilisé est plus loin du centre de gravité de la culture du laboratoire.

Une petite déclaration de politique scientifique.

- L'équipe AMACC visera à rester, comme l'équipe qui lui a donné naissance, une des équipes provinciales bien visibles parmi les équipes françaises de même thématique. Elle continuera à jouer fortement le jeu national, en ayant une politique active dans la structure du GdR IM. Elle continuera à s'investir dans les appels à projets ANR, et elle y espère la même réussite qu'auparavant.

- Elle continuera à animer, comme elle le fait maintenant, les séminaires hebdomadaires d'Algorithmique (en liaison avec l'équipe CODAG) et de Cryptologie (en liaison avec le laboratoire LMNO et l'équipe d'Orange Labs). Parallèlement, elle cherchera à recréer un groupe de travail plus proprement AMACC, peut-être deux fois par mois, qui pourra être un lieu d'expression plus informel pour ses membres, notamment ses doctorants. Ce groupe de travail peut être un instrument important pour renforcer les arêtes du triangle AMACC.

- Elle continuera à mener une politique ambitieuse de publications en revues, et accroîtra son ambition sur les conférences internationales de très bon niveau.

- Elle mènera une politique incitative en direction des membres du collègue B de l'équipe pour qu'ils préparent et passent leur HDR.

1.1.5 Risques et opportunités.

Risques. Avec le départ du thème T4, qui comprenait trois membres du collègue A, la nouvelle équipe AMACC voit son encadrement en membres du collègue A diminuer fortement, et tomber à deux unités, qui seront de plus bien proches de la retraite en 2015. L'un des deux est une DR CNRS, dont le remplacement est aussi moins automatique. De plus, même si le thème T2 peut profiter de l'encadrement administratif de membres A de l'équipe, il ne possède aucun vrai encadrement scientifique de rang A ; il faut donc, pour T2, de toute urgence, un recrutement en collègue A : cela doit être une priorité du laboratoire.

Un certain isolement culturel de l'équipe affaiblit son influence sur l'architecture des enseignements. Ce phénomène s'est accentué dans les nouvelles habilitations du Master. Il s'ensuit un recrutement local d'étudiants un peu difficile en M2 Recherche, notamment pour les thèmes T1 et T3, où le bagage des étudiants doit comporter une base importante de mathématiques. L'équipe doit compter sur son insertion nationale, et sur sa participation à d'autres M2 Recherche pour le recrutement de doctorants. Le recrutement pourrait sembler plus facile dans les thématiques de T2, qui exercent un grand attrait chez les étudiants. Mais, il faut, là aussi, un solide bagage mathématique, et, en plus, le thème T2 n'a pas de membre habilité! ...

Opportunités. Nous n'avons pas souhaité le départ de nos membres T4. Maintenant, nous pensons que la nouvelle configuration du laboratoire peut s'avérer très positive, en créant deux équipes AMACC et CODAG, chacune bien soudée autour de ses thèmes propres, dans une ambiance d'harmonie et de complémentarité.

1.2 Thème 1 : Modèles de calcul, logique et complexité.

Projet global. Les recherches de ce thème cherchent à établir une relation étroite entre la complexité algorithmique et, d'une part les modèles de calcul séquentiel (modèle RAM) et parallèle (modèle des automates cellulaires), d'autre part la logique (complexité descriptive).

Tout récemment, nous avons commencé à faire converger les recherches sur ces deux modèles de conception orthogonale, en mettant en évidence des caractérisations logiques du temps linéaire des AC, caractérisations très comparables à celles obtenues auparavant pour les classes de complexité des RAM. Ces résultats nous encouragent à développer une algorithmique en temps linéaire des AC et à la comparer avec l'algorithmique classique, riche et bien développée, du temps linéaire dans le modèle RAM.

Complexité optimale des requêtes (Johann Brault-Baron, Étienne Grandjean, Ionona Ranaivoson).

Dans la continuité des résultats déjà obtenus sur la complexité de l'énumération des requêtes conjonctives acycliques, on est en train d'établir une classification assez complète de la complexité des requêtes conjonctives dans les bases de données, en ce qui concerne les problèmes d'existence de solutions, leur comptage, leur énumération, enfin la recherche de la j -ème solution. Cette classification s'appuie sur des caractérisations de l'hypergraphe de la requête conjonctive et établit que la complexité du problème (de comptage, d'énumération, etc.) est la meilleure (c'est-à-dire, est en temps linéaire, à délai constant, etc.) dans le cas où l'hypergraphe associé est acyclique ou appartient à une sous-classe bien identifiée d'hypergraphes acycliques (sous l'hypothèse que l'hypergraphe est "conforme"). On généralise aussi cette classification aux disjonctions de requêtes conjonctives, c'est-à-dire à toutes les requêtes SQL de base. Ces résultats sont clairement reliés aux recherches menées dans la future équipe CODAG sur l'analyse de la structure des problèmes de satisfaction de contraintes (CSP) et celle des requêtes dans les bases de données ; comme le fait cette équipe, nous mettons en oeuvre pour cela les mêmes outils combinatoires : les graphes ou hypergraphes, associés aux CSP ou aux requêtes, enfin leurs décompositions arborescentes.

Caractérisations logiques de la complexité en temps linéaire des automates cellulaires et des langages d'images reconnaissables (Étienne Grandjean, Gaétan Richard).

Cette recherche tente de donner les premiers résultats de complexité descriptive quant au modèle classique des automates cellulaires (AC). Répondant à une question de J. Mazoyer (2000), le groupe caennais a obtenu récemment, en collaboration avec F. Olive du LIF, une caractérisation logique (en logique existentielle du second ordre) de la classe des problèmes calculables en temps linéaire sur un AC *non déterministe*, ceci pour toute dimension de l'automate. On établit aussi la robustesse de cette caractérisation et donc du temps linéaire des AC en démontrant que nombre de variantes syntaxiques de cette logique définissent la même classe de complexité. Enfin, on montre qu'une autre variante de cette logique définit la classe des langages d'images *reconnaissables*, extension classique en dimension 2 (ou plus) de la classe des langages rationnels sur les mots, et, au-dessus de la classe des langages d'images reconnaissables, on définit une hiérarchie de langages d'images dont on démontre qu'elle est stricte, au moins au premier niveau.

Au delà de ces résultats qui restent à affiner sur quelques points, notre objectif est la question importante d'exhiber une restriction naturelle (avec des formules de Horn) de la logique existentielle du second ordre qui caractériserait la classe des problèmes décidables en temps linéaire sur AC *déterministe*. Un second objectif lié et essentiel est de développer des procédures génériques calculables en temps linéaire sur AC et ainsi de déterminer des classes de problèmes linéaires sur AC. Ceci donnerait quelques bases claires pour comparer la complexité séquentielle des RAM et celle parallèle des AC, problème largement ouvert.

Algorithmique sur automates cellulaires (Gaétan Richard, Véronique Terrier).

Pour évaluer les capacités de calcul des AC, notre approche s'appuie principalement sur l'étude de ses classes de complexité et le développement d'algorithmes performants pour résoudre des problèmes spécifiques.

Quand on passe d'un réseau d'automates de dimension 1 à un réseau de dimension supérieure et que l'on augmente aussi la dimension des données, tout devient plus délicat. Notamment les interactions locales définies par le voisinage deviennent plus complexes. En particulier, dès la dimension 2, on ne connaît pas, pour l'instant, de résultat d'accélération linéaire qui s'applique quelque soit le voisinage. Sur ce sujet, une des pistes actuellement explorée est un nouvel algorithme d'accélération que nous avons établi pour la dimension 1 et que nous cherchons à généraliser à la dimension supérieure.

Sur un plan plus de nature constructive, le point de vue problèmes de calcul (contrairement aux problèmes de décision) n'est guère développé. Dans cette perspective, il s'agirait en s'appuyant en particulier sur les résultats obtenus par É. Grandjean relatifs à la complexité linéaire des machines RAM, de concevoir des algorithmes adaptés pour des problèmes caractéristiques ou de montrer leur difficulté.

On manque aussi d'exemples spécifiques qui illustrent les possibilités algorithmiques propres à la dimension 2. Et nous souhaiterions également étudier la comparaison des AC en dimension 2 avec les différents modèles déterministes qui généralisent en dimension 2 la classe des langages rationnels. Un point déjà partagé entre les AC et ces modèles, est que le passage à cette dimension 2 soulève de nombreuses difficultés.

Dans toutes les problématiques précédentes, l'objectif est de mieux comprendre la notion de calcul

parallèle au travers du modèle des automates cellulaires. Chaque pas dans ce domaine nécessite des outils adaptés pour trouver le bon point de vue. De ce fait, il sous-entend la poursuite du travail développé par G. Richard dans sa thèse sur les concepts formels de particules et collisions.

Applications des techniques fonctionnelles paresseuses (Jerzy Karczmarczuk).

Dans l'approche fonctionnelle des systèmes quantiques, après un premier travail, où nous avons construit un cadre de représentation des états et des opérateurs physiques, nous voulons maintenant analyser deux autres aspects du formalisme : l'universalité des constructions des systèmes composites par des produits tensoriels des états individuels (ce qui pour les systèmes infinis est très difficile), et exploiter le contenu "géométrique" de l'évolution quantique (l'application des opérateurs linéaires aux états), à l'aide de la logique linéaire de Girard (le mot *linéarité* possède ici un double sens, et ceci n'est pas accidentel).

Dans le prolongement de nos travaux sur les algorithmes de *différentiation automatique*, nous analysons à présent d'autres systèmes, où de tels paradigmes s'avèrent intéressants. En particulier, les similitudes avec les algorithmes d'apprentissage des réseaux neuronaux par rétropropagation des erreurs est un des chemins empruntés.

1.3 Thème 2 : Protection de l'information, codage et cryptographie.

Projet global. Comme l'introduction l'a déjà mentionné, la géométrie de ce thème se modifie : il intègre à la fois de nouveaux membres et de nouvelles problématiques de sécurité. Ces dernières se situent dans l'*internet des objets*, dans l'embarqué, dans le document, et s'inscrivent naturellement dans l'environnement du thème. Les thématiques plus historiques seront bien évidemment prolongées, en particulier dans les domaines de la conception de protocoles en protection de l'information (cryptologie, tatouage) et de l'étude des algorithmes et objets mathématiques sous-jacents.

Conception de protocoles de protection de l'information (A. Akhavi, M. Georgieva, F. Laguillaumie, J.-M. Le Bars, L. Lhote, J. Madelaine, B. Vallée).

Conception de systèmes cryptographiques. Nous étudions tout particulièrement les mécanismes de chiffrement dont les propriétés particulières, permettent de respecter des contraintes d'anonymat. Les chiffrements de différents types (homomorphe, basé sur l'identité, sur des attributs, supportant le re-chiffrement,...) sont des primitives qui ne sont pas toujours conçues de façon efficace. Leur sécurité, comme celle des schémas de signatures, doit également être étudiée dans un contexte multi-utilisateur (et ce n'est pas toujours le cas aujourd'hui). En particulier, nous nous intéresserons plus précisément à la notion de preuve de possession, nécessaire pour enregistrer une clé publique, dont les constructions actuelles ne sont pas satisfaisantes.

Même si la plupart des systèmes cryptographiques que nous étudions reposent sur des objets de la théorie des nombres, nous songeons également à utiliser des objets des mathématiques discrètes, et certaines propriétés aléatoires des graphes, pour définir des schémas d'identification.

Tatouage. Pendant la thèse de C. Bazin [4], nous avons étudié le tatouage de documents structurés, qui doit préserver des contraintes, et nous avons mené en particulier beaucoup d'expérimentations sur les documents géographiques. Cela nous a permis de mieux cerner un domaine encore peu développé et formalisé. Nous avons proposé un schéma général, qui sépare clairement une partie générique commune à tous les documents, d'une partie spécifique à un certain type de document. Cela montre qu'il est indispensable d'avoir une bonne connaissance de la problématique générale du tatouage, ainsi qu'une bonne expertise du type des documents à tatouer. Dans le futur, nous étudierons l'efficacité du schéma sur d'autres types de documents. En particulier, pour les documents textuels, la propriété locale doit clairement reposer sur des concepts liés à la linguistique et au traitement automatique des langues.

Cryptologie basée sur les réseaux euclidiens. Le chiffrement est une primitive cryptographique essentielle, qui comme la signature, peut être modifiée pour assurer une propriété de sécurité particulière. Récemment, C. Gentry a levé un solide verrou scientifique à STOC 2009, en proposant le premier chiffrement *complètement homomorphe*. La sécurité de ce système repose sur la difficulté de problèmes de réseaux

euclidiens. L'étude précise de la sécurité de ce système, et l'amélioration de ses performances, constituent des challenges particulièrement intéressants, qui motivent ce projet.

Dans la même veine, O. Regev a proposé à STOC 2005 un algorithme de chiffrement dont la sécurité repose sur un autre problème des réseaux euclidiens, le problème « learning with errors ». Nous cherchons à adopter cet algorithme pour obtenir des propriétés particulières : il peut être possible, par exemple, d'utiliser des techniques de délégation de base pour obtenir un chiffrement sûr basé sur les attributs.

Algorithmes et objets mathématiques sous-jacents (J.-M. Le Bars, F. Laguillaumie, L. Lhote, A. Otmani).

Énumération, codage énumératif et génération de fonctions booléennes. Dans la méthode par décomposition récursive de classes, que nous avons utilisée pour les fonctions 1-résilientes [12, 13], il faut stocker la cardinalité des classes (phase d'énumération), puis utiliser ces cardinalités pour effectuer le codage énumératif, qui est ensuite utilisé pour la génération aléatoire. Nous allons reprendre cette méthodologie novatrice pour étudier d'autres propriétés cryptographiques, comme l'immunité algébrique et la non-linéarité (travail commun avec Alfredo Viola, Antoine Genitrini et Sihem Mesnager).

Cryptanalyse algébrique. L'analyse théorique de la complexité de l'attaque algébrique est un axe important notamment pour obtenir des critères pour évaluer la sécurité de futures nouvelles variantes du système de McEliece. Pour cela, il faut étudier plus précisément les algorithmes génériques de résolution de systèmes algébriques, afin de mieux les adapter à notre contexte. Ce travail, actuellement en cours, vise en particulier à exploiter la forme très particulière des systèmes que nous rencontrons dans le cas du chiffrement de McEliece.

Nous souhaitons également étendre la zone de validité de l'algorithme de distinction d'un code de Goppa d'un code aléatoire. En effet, à l'heure actuelle, notre méthode ne s'applique qu'à des codes ayant des rendements proches de 1. Ce type de code se rencontre essentiellement dans le schéma de signatures CFS, mais, à l'inverse, les codes proposés dans le cadre de fonctions de chiffrement ont des rendements bien inférieurs à 1.

Enfin, dans un dernier volet plus ambitieux, à plus long terme, nous chercherons à améliorer nos méthodes d'étude des systèmes fondés sur la théorie des codes, afin d'obtenir une première cryptanalyse totale de ces systèmes, notamment du schéma CFS. Nous souhaitons en effet utiliser toutes les avancées, tant théoriques que pratiques, dans la résolution de systèmes algébriques, pour les appliquer dans les attaques des systèmes fondés sur les codes.

Théorie algorithmique des nombres, cryptanalyse. Sur l'algorithme de factorisation des entiers de la forme pq^2 [5], de nombreuses questions restent ouvertes : en particulier, cet algorithme arrive parfois à factoriser des entiers, même s'ils ne vérifient pas la condition arithmétique qui rend l'algorithme efficace. Par ailleurs, nous pensons que cette condition arithmétique sur p peut également permettre d'obtenir un analogue de l'algorithme ρ de Pollard dans l'infrastructure du groupe des classes. De façon générale, est-il vrai que les entiers de cette forme particulière, que l'on rencontre fréquemment en cryptographie, soient plus faciles à factoriser ? Les techniques particulières développées dans le cadre de ces entiers peuvent-elles s'adapter à des entiers de type RSA ?

Les couplages sur les courbes elliptiques sont un outil toujours aussi intéressant pour la conception système cryptographiques. À l'heure actuelle, les couplages commencent à intéresser les fabricants de cartes à puce, qui veulent savoir quelles courbes permettent d'implanter les couplages, sur quel corps de définition, ... L'état de l'art ne permet pas encore de répondre à ces questions. Éventuellement, chaque application pourrait nécessiter une courbe adaptée. Ce type d'étude continue à faire l'objet de collaborations avec nos collègues mathématiciens.

Algorithmique des réseaux euclidiens. De façon générale, l'algorithmique de la réduction des réseaux euclidiens est une aire de jeux appréciée des cryptologues, qui reste néanmoins mal connue. C'est pour cela que l'étude de l'algorithme LLL est très utile. En collaboration avec le Thème 3 de l'équipe, dans le cadre du projet ANR *Lareda*, nous cherchons à mieux analyser le comportement de cet algorithme dans la vie « réelle » de ses applications, en particulier en étudiant des versions simplifiées modélisables par des tas de sable [15]. Cette étude particulière, sur des réseaux issus de la cryptographie, peut donner de bonnes intuitions de la difficulté du problème général de la réduction.

Sécurité des systèmes embarqués et de l'internet des objets (F. Laguillaumie, J. Madelaine, J. Saquet).

L'Internet des Objets (IoT pour l'anglais « Internet of Things ») permet le suivi des biens de façon à permettre la traçabilité ou la détection de contrefaçons. On munit chaque objet d'une étiquette porteuse d'un code unique et lisible, grâce à une technologie sans contact (comme par exemple le RFID « Radio Frequency Identification »), et, à chaque fois qu'un lecteur détecte une étiquette et, par là même, l'objet porteur, on construit un évènement qu'on stocke dans un système d'information distribué. Nous avons développé en partenariat avec Orange Labs un tel système, en suivant les recommandations de EPCglobal : c'est la plateforme IoTa. Nous chercherons à concevoir et à étudier des protocoles cryptographiques légers et fiables, qui permettent par exemple de tracer des objets déployés à grande échelle, de façon sécurisée, tout en respectant la vie privée des utilisateurs. Les questions de sécurité se posent d'abord au niveau de la lecture de l'étiquette et seules des techniques de cryptologie bas-coût sont utilisables à ce niveau, compte-tenu de la faible capacité de calcul des étiquettes. Ils se posent ensuite au niveau des échanges entre composants. Même si ces questions ne posent pas de réels problèmes théoriques, la mise en œuvre de la sécurité de bout en bout dans la plateforme IoTa serait une première. Au niveau de la confidentialité, l'existence de services de référencement (à la DNS) et de découverte (type moteur de recherche) peuvent de facto délivrer plus d'information que ne le voudraient les règles de contrôle d'accès définies au niveau du stockage des évènements. Le projet TACITES, qui reprend ces thématiques de recherche, a été déposé au FUI 10.

1.4 Thème 3 : Structures aléatoires et analyse d'algorithmes.

Projet global. Comme dans la période précédente, les recherches de ce thème se déclinent selon trois volets :

- un premier volet plus méthodologique, qui cherche à modéliser le concept de source, et les propriétés qui seront essentielles dans l'analyse des structures de données construites sur ces sources ;
- un second volet, dédié à l'analyse probabiliste des algorithmes du texte ;
- un troisième et dernier volet, qui, lui, évolue. Dédié auparavant à l'arithmétique, il est maintenant dédié, plus généralement, à l'analyse d'objets qui interviennent en cryptographie.

Bien sûr, ces volets ne sont pas étanches. Un fil conducteur important (mais non omni-présent) reste l'analyse dynamique, spécialité locale qui mélange méthodes d'analyse en moyenne et méthodes issues de systèmes dynamiques.

Sources et théorie de l'information (J. Clément, L. Lhote, M. Roux, B. Vallée). Il s'agit d'une activité de modélisation autour du concept de source de la théorie de l'information.

Les régions sans pôles des séries génératrices des sources. L'équipe a défini un modèle de source très général. Une source sur l'alphabet Σ est complètement déterminée par la famille des probabilités fondamentales p_w , quand w décrit Σ^* . On lui associe la série de Dirichlet $\Lambda(s)$ de terme général p_w^s , qui n'a pas de pôles pour $\Re s > 1$. La géométrie des pôles de $\Lambda(s)$ autour de la droite $\Re s = 1$ est essentielle pour décrire l'asymptotique fine des structures de données et des algorithmes travaillant sur les mots de la source. La thèse de M. Roux est centrée sur la géométrie de ces pôles et un premier résultat a déjà été obtenu dans [11], dans le cas d'une source sans mémoire sur un alphabet fini, où l'on montre que la forme de la région sans pôles dépend des propriétés d'approximabilité de la famille de probabilités (p_i) . Il s'agit de relier, plus généralement, la forme de la région sans pôles de $\Lambda(s)$ et les propriétés arithmétiques de la source. Dans le cas des sources associées à un système dynamique, il faut étendre à l'opérateur « sécant » les propriétés établies par Dolgopyat, Naud, Melbourne pour l'opérateur « tangent ». Une première étape a déjà été franchie dans [6].

Entropie. Ces dernières années, l'équipe a développé une collaboration avec les mathématiciens de Caen et organise un groupe de travail (pas encore très régulier) autour des problématiques sur les mots, les sources ou les entropies. Cette collaboration s'est traduite par un premier résultat sur le calcul et l'estimation de taux d'entropie généralisés [9]. Nous souhaitons conforter cette collaboration en « régularisant » le groupe de travail, et en généralisant le premier résultat obtenu à d'autres modèles de sources.

Codes bifixes. La classe des codes bifixes a des propriétés très intéressantes pour la synchronisation, car les mots de ces codes peuvent être décodés instantanément (i.e., dès leur lecture) quel que soit le sens de leur lecture. Depuis qu'elle a été introduite, cette classe se révèle bien plus difficile à étudier que celle des codes préfixes, et donne lieu à des conjectures non élucidées. Dans ce cas, on ne connaît par exemple pas d'analogue du célèbre algorithme de Huffman. Après l'étude fructueuse que nous avons faite sur les codes préfixes optimaux pour des sources sans mémoire à distribution géométrique [2, 3], nous projetons d'étudier cette classe de codes bifixes dans le même cadre de sources (J. Clément, dans une collaboration avec Frédérique Bassino, Gadiel Seroussi et Alfredo Viola).

Algorithmique du texte : recherche de motifs, algorithmes de tri et de recherche (J. Clément, L. Lhote, B. Vallée).

Recherche de motifs. Nous continuons à approfondir l'analyse probabiliste des motifs généraux dans les textes, ou des traverses minimales dans les hypergraphes.

- Dans le cas où un motif X est un ensemble de mots finis, le nombre d'occurrences N_X de X dans un texte est (par définition) la somme des nombres d'occurrences de chaque mot. On considère deux motifs X et Y , et on s'intéresse aux propriétés probabilistes du couple (N_X, N_Y) . On cherche à caractériser les couples (X, Y) de motifs pour lesquels le couple (N_X, N_Y) a un comportement asymptotiquement normal, et on souhaite simplifier des anciens critères donnés par Bender et Kochmann qui garantissent une loi normale conjointe.

- Le comportement de nombreux algorithmes de recherche d'un motif n'est pas lié au motif lui-même mais seulement à sa structure d'autocorrélation (recouvrements possibles du motif avec lui-même). C'est le cas par exemple pour le célèbre algorithme de Knuth-Morris-Pratt, qui dépend uniquement de la table des bords (ou de manière équivalente de la table des préfixes). Si l'on veut faire une analyse probabiliste d'un algorithme de ce type, il est naturel de générer la table des préfixes plutôt que le motif lui-même. Cette génération aléatoire dépend donc de la taille des classes d'équivalence pour la relation « avoir même table », et il faut donc analyser ces différentes tailles possibles (Julien Clément en collaboration avec Maxime Crochemore, Cyril Nicaud et Giuseppina Rindone).

- Il n'existe pas aujourd'hui d'algorithme polynomial capable d'extraire toutes les traverses minimales d'un hypergraphe. Nous avons montré, dans un travail en cours de soumission [14] que l'algorithme **MTMiner**, proposé par l'équipe DoDoLa (et donc la nouvelle équipe CODAG), est presque sûrement *output-polynomial* dans un modèle aléatoire à la Erdős et Rényi. Nous souhaitons étendre ce point de vue original à l'analyse en moyenne d'autres algorithmes d'extraction de traverses minimales, dans d'autres modèles aléatoires.

Analyse réaliste des algorithmes de tri et de recherche. Il s'agit de deux types d'algorithmes : les algorithmes de texte, qui utilisent des arbres dictionnaire, ou les algorithmes généraux, qu'on étudie quand les clés sont vues comme des mots.

- L'équipe a déjà analysé les algorithmes de recherche fondés sur un arbre-dictionnaire (le trie), quand les mots de la source sont émis par une source générale. Elle a aussi pris en compte des propriétés fines liées à l'implémentation réaliste de telles structures. Il s'agit maintenant pour J. Clément de rédiger un chapitre sur ces analyses probabilistes fines des arbres-dictionnaire, dans un livre [8] (prévu pour 2011) en collaboration avec Brigitte Chauvin et Danièle Gardy.

- L'article [10] a ouvert la voie à un nouveau champ de recherche. Il s'agit de revisiter l'analyse en moyenne des principaux algorithmes de tri et de recherche, lorsque les clés sont des mots produits par une source générale et le coût de la comparaison entre clés est le nombre de symboles utilisés pour comparer les mots. L'article a déjà analysé les algorithmes QuickSort et QuickSelect de ce point de vue d'analyse réaliste. L'équipe vient d'obtenir un financement de thèse CNRS/Région pour étendre ce point de vue aux principaux algorithmes de tri et de recherche (début en septembre 2010). C'est aussi un travail qui est une des principales pistes du projet ANR **Magnum** qui vient d'être accepté, et auquel participe B. Vallée, en collaboration avec le LIP6, le LIAFA et le LIX.

Analyse probabiliste d'objets « cryptographiques ». Ces recherches sont de fait toutes motivées par la cryptologie, avec des degrés de motivation divers et se font en forte interaction avec le Thème 2. Cet axe de recherche est original et porteur, car, en général, les cryptologues se préoccupent assez peu d'analyse probabiliste, et les analyseurs en moyenne, assez peu de cryptologie...

Fonctions booléennes (L. Lhote). J.-M. Le Bars et A. Viola ont proposé une méthode pour énumérer les fonctions 1-résilientes [12, 13]. Ils veulent maintenant générer des fonctions booléennes ayant d'autres propriétés cryptographiques (et pourquoi pas, plusieurs propriétés simultanées). Pour le thème 3, il s'agit d'étudier la combinatoire asymptotique de ces fonctions, en utilisant des méthodes de col avec un très grand nombre de variables. Plus généralement, nous voudrions adapter les méthodes de col afin d'en faire des boîtes à outils à l'image des dictionnaires de la combinatoire analytique. D'autres problèmes combinatoires comme les phénomènes de seuil des problèmes SAT pourraient en bénéficier.

Algorithmique des réseaux euclidiens (A. Akhavi, J. Clément, M. Georgieva, L. Lhote, B. Vallée). Ces recherches s'effectuent dans le cadre du projet ANR **Lareda**

- L'article [15] a proposé, sous une hypothèse de régularité, une modélisation simplifiée de l'algorithme LLL par les tas de sable. Dans le cadre de la thèse de M. Georgieva, il est projeté de comparer plus précisément ces modèles simplifiés à l'algorithme réel, et d'en déduire, éventuellement, des heuristiques semi-prouvées d'accélération du « véritable » algorithme LLL.

- Dans le cas où la dimension du réseau est beaucoup plus petite que la dimension ambiante, l'article [1] montre qu'il est très efficace d'utiliser d'abord une projection aléatoire bien choisie. Il reste à définir les distributions de projections aléatoires qui conviennent.

- L'analyse de l'algorithme LLL pour la dimension $n = 3$ pourrait être un test pour la méthode symbolique introduite par A. Akhavi, déjà appliquée avec succès à certains algorithmes de tri et à l'algorithme de réduction de Gauss (LLL pour $n = 2$).

Etude probabiliste des générateurs pseudo-aléatoires. L'article [7] a ouvert la voie de l'étude probabiliste fine d'une classe particulière de générateurs pseudo-aléatoires, les progressions arithmétiques modulaires. Nous voulons généraliser cette étude dans plusieurs directions : nouvelles mesures de « randomness », nouvelles classes de générateurs pseudo-aléatoires, extension des méthodes dynamiques. Ce champ d'étude, très large, se fera en collaboration avec E. Cesaratto (Buenos-Aires) et J. Bourdon (Lina, Nantes).

Etude probabiliste de l'algorithme de Berlekamp-Massey. Cet algorithme est à la fois un outil de cryptanalyse et une brique essentielle pour le décodage de codes correcteurs. Il existe une version de cet algorithme qui est proche d'une version de l'algorithme d'Euclide (dit « étendu et interrompu »). Dans le cadre d'un projet de Master, on a effectué l'analyse en moyenne de cette version de l'algorithme d'Euclide, et on doit maintenant transposer cette analyse au véritable algorithme de Berlekamp-Massey.

1.5 Les projets de l'équipe en cours en 2011.

Projet **LAREDA** (2007–2011) du programme Blanc de l'ANR.

Lattice Reduction Algorithms : Dynamics, Probabilities, Experiments, Applications.

Coordinateur : GREYC – Autres partenaires : IMB (Dijon), IRISA (Rennes), LIP (Lyon), LIRMM (Montpellier).

Participants de l'équipe : A. Akhavi, J. Clément, F. Laguillaumie, L. Lhote, B. Vallée.

Objectif : analyse probabiliste des algorithmes de réduction de réseaux euclidiens et applications, notamment en cryptologie.

Projet **ENUM** (2007-2011) du programme Blanc de l'ANR.

Algorithms and complexity for answer enumeration.

Coordinateur : ELM (Paris 7) – Autres partenaires : GREYC, LIFL (Lille), LIF (Marseille).

Participant de l'équipe : E. Grandjean.

Objectif : étude des problèmes d'énumération, du point de vue des classes de complexité algorithmique comme de celui des applications : algorithmes d'énumération des solutions pour les requêtes en logique, dans les bases de données (XML, etc.), et pour les CSP.

Projet **PACE** (2007–2011) du Programme **Telecom** de l'ANR.

Pairings and Advances in Cryptology for e-cash.

Coordinateur : Orange Labs – Autres partenaires : LIENS (ENS Ulm), LIX(Ecole Polytechnique), GREYC, ST Ericsson, Cryptolog International, Gemalto.

Participant de l'équipe : F. Laguillaumie

Objectif : étude des systèmes de monnaie électronique qui visent à émuler électroniquement la monnaie fiduciaire en garantissant en particulier l'anonymat quasi parfait des utilisateurs au moment des dépenses.

Projet **Boole** (2009–2012) du programme Blanc de l'ANR.

Coordinateur : PRISM (Versailles)– Autres partenaires : GREYC, ENS Paris, INRIA Rocquencourt, LIF (Marseille).

Participants de l'équipe : A. Akhavi, J. Clément, J-M Le Bars, L. Lhote, B. Vallée.

Objectif : quantification des propriétés de structures booléennes avec des méthodes de combinatoire analytique ou des méthodes probabilistes.

Projet **EMC** (2009-2013) du programme Blanc de l'ANR.

Emergence dans les modèles de calcul

Coordinateur : LIF (Marseille) – Autres partenaires : I3S (Nice).

Participante de l'équipe : V. Terrier

Objectif : étude des propriétés intrinsèques aux systèmes complexes et de leurs aspects calculatoires, ciblée sur les contributions que l'informatique théorique peut apporter à ce domaine.

Projet **MAGNUM** (2010-2014) du programme Blanc de l'ANR.

Méthodes algorithmiques de génération aléatoire non uniforme. modèles et applications

Coordinateur : LIP6 (Paris 6) – Autres partenaires : LIAFA (Paris 7), LIX (Ecole Polytechnique).

Participante de l'équipe : B. Vallée

Projet **Wings** (2009–2011) du programme « Réseaux du futur et services »

Widening Interoperability for Networking Global Supply Chains.

Coordinateur : GS1 France – Autres partenaires : GREYC, Inria, Université Paris 6, Afnic, Orange Labs.

Participants de l'équipe : J. Madelaine, J. Saquet.

Objectif : fédérer les partenaires européens autour d'une structure de gestion de l' »Internet des objets » (Internet of Things ou IOT); mettre en commun les programmes nécessaires à la création d'une architecture répartie pour l'IOT en Europe.

Bibliographie

- [1] Ali Akhavi and Damien Stehlé. Speeding-Up Lattice Reduction with Random Projections (Extended Abstract). In Eduardo Sany Laber, Claudson F. Bornstein, Loana Tito Nogueira, and Luerbio Faria, editors, *LATIN 2008 : Theoretical Informatics, 8th Latin American Symposium*, volume 4957 of *Lecture Notes in Computer Science*, pages 293–305, April 7-11, Búzios, Brazil, 2008. Springer.
- [2] Frédérique Bassino, Julien Clément, Gadiel Seroussi, and Alfredo Viola. Optimal prefix codes for pairs of geometrically-distributed random variables. In *2006 IEEE International Symposium on Information Theory*, page 5 pages, Seattle, WA, USA, 2006.
- [3] Frédérique Bassino, Julien Clément, Gadiel Seroussi, and Alfredo Viola. Optimal Prefix Codes for Some Families of Two-Dimensional Geometric Distributions. In *DCC*, pages 113–122, Salt Lake City, Utah, USA, 2006.
- [4] Cyril Bazin, Jean-Marie Le Bars and Jacques Madelaine. A Novel Framework For Watermarking : The Data-Abstracted Approach International Workshop on Security, IWSEC2008, Lecture Notes In Computer Science, pp 201-217 (2008)
- [5] Guilhem Castagnos, Antoine Joux, Fabien Laguillaumie, and Phong Nguyen. Factoring pq^2 with Quadratic Forms : Nice Cryptanalyses. In Mitsuru Matsui, editor, *ASIACRYPT 2009, 15th International Conference on the Theory and Application of Cryptology and Information Security*, volume 5912 of *Lecture Notes in Computer Science*, pages 469–486, Tokyo, Japan, Deember 2009. Springer.
- [6] Eda Cesaratto and Brigitte Vallée. Gaussian distribution of trie depth for dynamical sources. En cours de soumission (25 pages).
- [7] Eda Cesaratto, Alain Plagne and Brigitte Vallée. On the non-randomness of modular arithmetic progressions : a solution to a problem by V. I. Arnold. *Comptes-Rendus du Colloquium on Mathematics and Computer Science : Algorithms, Trees, Combinatorics and Probability*, pp 271–288, DMTCS, 2006.
- [8] Brigitte Chauvin, Julien Clément, Danièle Gardy. Arbres en algorithmique. *Collection « Mathématiques et Applications »* de la SMAI (Société des Mathématiques Appliquées et industrielles), Springer (prévu en 2011)
- [9] Gabriela Ciuperca, Valérie Girardin, and Loïck Lhote. Computation of Generalized Entropy Rates. Application and Estimation for Countable Markov chains. Rapport de recherche 2010-06, Laboratoire de Mathématiques Nicolas Oresme, Université de Caen, 10 pages, 2010.
- [10] Julien Clément, James Allen Fill, Philippe Flajolet, and Brigitte Vallée. The Number of Symbol Comparisons in QuickSort and QuickSelect. In S. Albers, editor, *ICALP 2009*, number 5555 (Part I) in LNCS, pages 750–763, 2009.
- [11] Philippe Flajolet, Mathieu Roux, and Brigitte Vallée. Digital trees and memoryless sources : from arithmetics to analysis. In *AofA10*, DMTCS, 2010. 12 pages, to appear.
- [12] Jean-Marie Le Bars and Alfredo Viola. Equivalence classes of boolean functions for first-order correlation. In *IEEE International Symposium on Information Theory, 2007. ISIT 2007*, pages 181–185, Nice, France, June 2007.
- [13] Jean-Marie Le Bars and Alfredo Viola. Equivalence classes of Boolean functions for first-order correlation. *IEEE Transactions on Information Theory*, 56(3) :1247–1261, Mars 2010.
- [14] Loïck Lhote. Mining minimal transversals in large random hypergraphs is with high probability output-polynomial. En cours de soumission (20 pages).

- [15] Manfred Madritsch and Brigitte Vallée. Modelling the LLL algorithm by sandpiles. In *Comptes Rendus de la Conférence LATIN 2010*, pages 267–281, Oaxaca, Mexico, 2010. LNCS 6034.

2

Équipe COnstraints, DAta mining and Graphs

2.1	Introduction	29
2.1.1	Cœurs de métiers	31
2.1.2	Collaborations existantes	31
2.2	Projet de recherche	32
2.2.1	Thématiques transversales	33
2.2.2	Projets propres aux "Cœurs de métiers"	35
2.3	Vie de l'équipe	36

Equipe CoDaG

COntraintes, DAta mining et Graphes

Responsable : Arnaud Lallouet

L'équipe CoDaG, nouvellement créée dans le cadre du quadriennal 2012-2015, provient de la fusion du Thème 4 "Contraintes et Graphes" de l'ancienne équipe Algorithmique et de la partie "Fouille de données" de l'ancienne équipe "Données, Document, Langue". Cette évolution s'est préparée au cours du quadriennal 2008-11 à la fois par des collaborations qui se sont intensifiées au cours de cette période et par une réflexion sur les opportunités scientifiques de nos activités. Les différents éléments ayant contribué à la création de l'équipe sont décrits aux sections 2.1.1 et 2.1.2.

L'équipe CoDaG se propose de rassembler de façon cohérente des compétences verticales en Aide à la Décision en intégrant les aspects Fouille de données et Optimisation. Le rassemblement de compétences en graphes, contraintes et fouille de données ainsi qu'une expérience sur des problématiques de recherche nécessitant la coopération de techniques venant de plusieurs de ces domaines forment une des originalités de l'équipe. Cette genèse permet à l'équipe de se positionner sur des thématiques émergentes et prometteuses telles que par exemple les méthodes fondées sur les contraintes pour la fouille de données, l'hybridation fouille/traitement automatique des langues (TAL) ou les recherches de sous-problèmes polynômiaux en contraintes.

L'équipe s'organise autour de trois "cœurs de métiers" et quatre "thématiques transversales".

Cœurs de métiers	{	CMG : Graphes, aspects algébriques et algorithmiques ; CMC : Contraintes, méthodes de résolution et d'optimisation ; CMF : Fouille, représentations condensées et synthèse.
Thématiques transversales	{	TT1 : Hybridation fouille de données et TAL ; TT2 : Fouille de graphes pour la chimie ; TT3 : Méthodes à base de contraintes pour la fouille de données ; TT4 : Graphes et contraintes.

2.1 Introduction

Contexte de création de l'équipe CoDaG. Forte de compétences en **graphes, contraintes et fouille de données**, l'équipe CoDaG a émergé informellement en 2009/2010 à travers un groupe de travail rassemblant des chercheurs des équipes Algorithmique et "Données, Document, Langue". L'équipe CoDaG se propose de développer des travaux suivant ces trois thématiques et en cherchant à renforcer les connexions entre celles-ci.

En particulier, de **fortes collaborations** existent entre la fouille de données et le TAL où l'on recherche une intégration fine de la modélisation linguistique aux algorithmes de fouille et entre les contraintes et la fouille de données, à travers la conception de contraintes globales dédiées fouille. Il est à noter que ces collaborations se sont fortement **intensifiées en 2009-2010** (18 publications communes). L'intérêt majeur de ces liens forts est qu'ils posent des problèmes nouveaux et qu'ils ne sont pas une application simple des techniques d'un domaine à un autre. Ces différents aspects sont développés dans les sections 2.1.1 et 2.1.2.

Interactions au sein du GREYC. Au sein du GREYC, les séminaires Algorithmique et I3 seront un cadre naturel pour continuer les échanges scientifiques avec les autres équipes issues de la réorganisation du GREYC. Les liens avec l'équipe DLU vont rester forts, la fouille de données textuelles formant un pont entre l'aspect méthodologique, relevant plus de l'équipe CODAG et, le volet plus applicatif, provenant pour cette activité, de l'équipe DLU. Les interactions avec AMACC sur l'étude de la complexité moyenne d'algorithmes de fouille de données seront poursuivies, tout comme l'étude de liens entre les notions de graphe et les problèmes de bases de données. La problématique commune de l'aide à la décision rapproche les équipes CODAG et MAD et Thierry Charnois va participer au co-encadrement avec Maroua Bouzid d'une thèse CIFRE avec EADS sur le raisonnement temporel et spatial. Une collaboration co-pilotée par

Bertrand Cuissart et Luc Brun est déjà prévue sur le développement et l'utilisation de méthodes fondées sur les graphes en chémoinformatique. Les deux équipes CODAG et Image partagent aussi des intérêts communs en apprentissage automatique et fouille de données. Un membre de l'équipe CODAG échange avec l'équipe Monétique & biométrie dans le cadre du projet ASAP.

Opportunités Le rapprochement de techniques algorithmiques liées à l'extraction d'information dans de gros volumes de données et la conceptualisation de ces techniques à travers les contraintes, les liens entre la modélisation linguistique et la découverte d'information, les graphes et contraintes ainsi que leurs usages dans de nombreux domaines comme la chimie et le texte sont très certainement des thèmes porteurs actuels. Au niveau national et international, il existe des laboratoires concurrents qui ont identifié les mêmes défis et souhaitent également les relever dans un cadre unifié. On peut citer le LIMOS (Projet Graphes, Algorithmes et Fouille de Données, Limoges), le LIFO (Contraintes et Apprentissage, Orléans), le LIRMM (Projet Coconut, Montpellier), le LIRIS (Département "Données, Connaissances, Services", Lyon) ou le projet CP4IM de l'Université de Leuven (Belgique). Il s'agit très certainement d'un environnement favorable pour assurer une visibilité à notre équipe car nous disposons d'une certaine ancienneté sur ce thème.

Risques Nous avons identifié deux points faibles sur lesquels nous allons travailler activement afin de ne pas pénaliser le lancement de l'équipe. Tout d'abord il s'agit du faible nombre de contrats et de financements qui peut à moyen terme ralentir l'obtention de nouveaux résultats. Il serait raisonnable de compléter les doctorants issus de la dotation du laboratoire par des doctorants et des chercheurs sur contrats. Ensuite le trop faible nombre de Maîtres de Conférences (3 (les deux MC à 50% étant comptés pour 1) pour la partie fouille, 1 pour la partie contraintes et 1 pour la partie graphes) ainsi que l'éloignement géographique de certains (3 MC en IUT dont 1 à Alençon) est un frein au développement de certaines thématiques. Face à cette situation difficile, il est souhaitable de recruter 3 MC : 1 en fouille, 1 en contraintes et 1 en graphes.

L'équipe comporte 9 membres permanents (deux membres sont à 50% avec l'équipe DLU) dont 4 collègue A (PR UFR Sciences) et 5 collègue B dont la plupart ont une activité d'enseignement délocalisée (1 MC UFR Sciences, 1 MC UFR LVE, 2 MC IUT Caen, 1 MC IUT Alençon) comme indiqué dans le tableau suivant :

Permanents			
Prénom et Nom	Fonction	Pourcentage	Provenance
Patrice Boizumault	PR UFR Sciences	50% (avec DLU)	Algo
Alain Bretto	PR UFR Sciences		Algo
Thierry Charnois	MC IUT Caen		DoDoLa
Bruno Crémilleux	PR UFR Sciences		DoDoLa
Bertrand Cuissart	MC IUT Alençon		Algo/DoDoLa
Jean-Jacques Hébrard	MC UFR Sciences		Algo
Arnaud Lallouet	PR UFR Sciences		Algo
Samir Loudni	MC IUT Caen		Algo
François Rioult	MC UFR Sciences	50% (avec DLU)	DoDoLa
Antoine Widlöcher	MC UFR LVE		DoDoLa

Non-permanents			
Prénom et Nom	Statut	Encadrant	Date
Leila Ben Othman	Doctorant	FR et B.Cré (co-tutelle Tunis)	10/2007
Ryan Bissell-Siders	Postdoc	B.Cui et B.Cré	03/2009
Xavier Dupont	Doctorant	AL	10/2010
Mathieu Fontaine	Doctorant	SL et PB	10/2009
Cerasela Jaulin	Doctorant	AB	10/2006
Mehdi Khiari	Doctorant	PB et B.Cré	10/2008
Jean-Philippe Métivier	Postdoc	PB et SL	10/2010
Guillaume Poezevara	Doctorant	B.Cui et B.Cré	10/2008
Yannick Silvestre	Doctorant	AB	10/2007

Avant de présenter les futurs travaux de l'équipe CoDaG, nous rappelons ceux réalisés par les différents membres dans leurs équipes respectives, montrant ainsi l'assise scientifique de notre projet.

2.1.1 Cœurs de métiers

CMG - Graphes.

PARTICIPANTS : A. Bretto, B. Cuissart, J.-J Hébrard, R. Bissell-Siders (post-doct), C. Jaulin, G. Poezevara et Y. Silvestre (doctorants).

La recherche concernant les graphes prend ici deux aspects : des études fondamentales dans le cadre de la théorie algébrique des graphes et des travaux plus appliqués de conception d’algorithmes. Dans le cadre de la théorie algébrique des graphes, nous avons développé la notion de G-graphes, qui sont une généralisation des cartes ou des dessins d’enfants. Ces travaux ont conduit à plusieurs applications : amélioration de la liste des graphes cubiques symétriques, la "Foster census", première liste de graphes cubiques semi-symétriques, ou graphes “diaboliques” utilisés en parallélisme et en automatique, étude de l’hamiltonicité des graphes de Cayley. Du côté de la conception d’algorithmes, nous nous intéressons aux méthodes d’appariement de graphes. Après avoir conçu un algorithme de calcul du plus grand sous-graphe commun à deux graphes, nous avons réalisé une méthode d’extraction des graphes fréquents émergents à partir d’une famille de graphes.

CMC - Contraintes.

PARTICIPANTS : P. Boizumault, A. Lallouet, S. Loudni, J.-P. Métivier (postdoc), X. Dupont, M. Fontaine et M. Khiari (doctorants).

Nous avons principalement étudié les mécanismes de recherche pour la satisfaction et l’optimisation, dans le cadre des *Constraint Satisfaction Problems* (CSP) valués (WCSP) et quantifiés (QCSP). Nous avons proposé une méthode de recherche hybride *anytime* et sa mise en oeuvre avec succès pour résoudre un problème de réservation en-ligne de connexions dans les réseaux ATM pour France Telecom R&D Lannion. Dans le cadre des WCSP et pour les recherches arborescentes, nous avons proposé de nouvelles heuristiques de sélection de valeur et de variable qui exploitent l’historique des solutions, la topologie du réseau et les coûts (Thèse de Nicolas Levasseur). Nous avons proposé de nouvelles sémantiques de relaxation de contraintes globales relaxées, ainsi que les tests de cohérence et algorithmes de filtrage associés ; puis nous avons appliqué ce cadre à la résolution de problèmes d’emplois du temps de personnel hospitalier (Thèse de Jean-Philippe Métivier). Enfin nous avons proposé un langage facilitant la modélisation et la résolution de CSP quantifiés en utilisant la quantification restreinte, ainsi qu’une extension aux problèmes d’optimisation. Ce cadre offre un langage pour modéliser les problèmes multi-niveaux étudiés en Recherche Opérationnelle et en théorie des jeux. Nous diffusons actuellement QeCode, le seul solveur de QCSP publiquement disponible.

CMF - Fouille de Données.

PARTICIPANTS : B. Crémilleux, B. Cuissart, F. Rioult, R. Bissell-Siders (post-doct), L. Ben Othman, M. Khiari et G. Poezevara (doctorants).

De nombreux travaux fondateurs en fouille de données ont porté sur l’extraction de motifs locaux, c’est-à-dire de régularités observées dans certaines parties des données. Initialement, le groupe s’est plus particulièrement intéressé à différents aspects de l’extraction et des usages de ces motifs. Il a développé une vision unifiée des usages multiples des représentations condensées de motifs ensemblistes issus de bases de données, afin de mieux mettre en évidence leurs apports et potentialités. Nous avons ainsi plus particulièrement acquis un savoir-faire dans les représentations condensées de motifs et la découverte de motifs sous contraintes. Tout en poursuivant sur cette voie générale de la “fouille de données orientée motifs”, nous nous intéressons plus particulièrement aux défis posés par l’abondance de résultats et la recherche de motifs de plus haut niveau et synthétiques. Pour cela, nous suivons une approche fondée sur les contraintes afin de définir de façon déclarative des motifs plus riches et de limiter le nombre de motifs produits. Une approche complémentaire est l’intégration de spécificités du domaine des données au cœur des méthodes de fouille, nous menons cette démarche notamment en TAL.

2.1.2 Collaborations existantes

Les liens entre graphes, contraintes et fouille de données restent jusqu’à aujourd’hui assez peu étudiés. Le GREYC présente l’originalité de rassembler des compétences dans ces domaines et les premiers résultats de nos collaborations montrent l’intérêt de renforcer de tels liens.

- Fouille de Graphes et d’Hypergraphes.

PARTICIPANTS : A. Bretto, B. Crémilleux, B. Cuissart, R. Bissell-Siders (post-doct), G. Poezevara (doctorant).

En établissant les liens entre l'extraction de motifs en fouille de données et le calcul des traverses minimales d'hypergraphes, Céline Hébert a proposé, dans sa thèse, une méthode de calcul de ces dernières particulièrement performante pour les hypergraphes denses. Ce travail s'est aussi avéré utile pour la visualisation de données. Dans le cadre plus général de la fouille de graphes, le groupe s'intéresse à tirer parti des avancées obtenues sur les isomorphismes de sous-graphes afin de les exploiter dans le contexte des derniers développements en fouille de données (représentations condensées, motifs contraints) par exemple pour la découverte de sous-graphes émergents. Un fil conducteur applicatif de ce travail est celui du traitement de l'information chimique, tel que la caractérisation de toxicophores, notamment dans le cadre du projet INNOTOX2.

- Contraintes et Fouille de Données.

PARTICIPANTS : P. Boizumault, B. Crémilleux, M. Khiari (doctorant).

Pouvoir définir des contraintes portant sur plusieurs motifs est un défi majeur en fouille de données pour produire les modèles demandés par les utilisateurs (e.g. classifieurs, découverte de sous-groupes). Il est clairement intéressant d'étudier les liens entre la découverte de motifs sous contraintes et les travaux menés en programmation par contraintes. Mais, la transposition sous forme d'un CSP d'un problème de fouille de données est loin d'être immédiate. Avec la thèse de Mehdi Khiari, nous contribuons à la découverte de motifs ensemblistes sous contraintes (cadre générique pour l'extraction de motifs locaux, représentations condensées pour des contraintes quelconques). Cette recherche est menée à travers le projet BINGO2 qui est porté par le GREYC.

- Fouille, Contraintes et TAL.

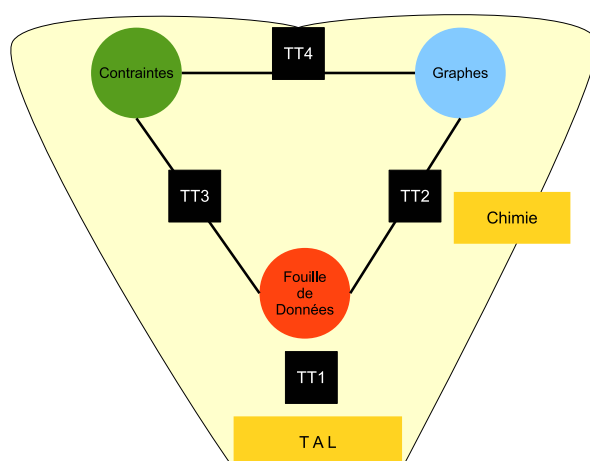
PARTICIPANTS : T. Charnois, B. Crémilleux, F. Rioult, A. Widlöcher.

Depuis plusieurs années, des collaborations entre les membres de CoDaG ont mis en évidence la pertinence de l'articulation profonde entre fouille, contraintes et TAL. Au sein des projets BINGO (ACI "masses de données" 2004-07), projet ANR ANNODIS (2008-10), la participation commune aux défis "fouille de texte" (DEFT 06, 07 et 08) ou à l'occasion des stages post-doctoraux de Marc Plantevit (2008-2009) et de Peggy Cellier (2009-2010), différents travaux ont déjà permis d'envisager une fouille de motifs en adéquation avec la matière linguistique, d'élaborer de nouvelles méthodes et de nouveaux motifs adaptés à cet objet d'étude (fouille récursive, motifs LSR ...) et d'intégrer aux processus de fouille des contraintes linguistiques permettant de garantir la pertinence des motifs découverts. Plus généralement, l'importance du paradigme des contraintes pour la modélisation des structures textuelles et pour leur exploration en fouille a constitué un important fil directeur. Ces différentes avancées ont donné lieu à de premiers résultats et à des publications.

2.2 Projet de recherche

Le projet de l'équipe repose sur un maintien des activités autour des cœurs de métier existants mais également sur une plus grande interaction. Celle-ci sera facilitée par l'intégration des membres au sein d'une même équipe et permettra de se doter d'outils pour faciliter la communication et les échanges entre des personnes qui font partie de communautés différentes.

Nous pensons que ces thématiques transversales sont une des originalités de l'équipe et notre projet se fonde à la fois sur celles-ci et nos cœurs de métier. L'activité scientifique menée dans ces thématiques transversales est liée à celle réalisée dans les cœurs de métier et ces différentes interactions sont schématisées par la figure suivante :



Nous débutons la présentation de notre projet scientifique par nos thématiques transversales puis nous donnons les perspectives de nos cœurs de métier sur lesquels s'appuient les thématiques transversales. Nous proposons ainsi une lecture croisée de notre projet.

2.2.1 Thématiques transversales

Au cours de la période du quadriennal précédent, de **nombreux points de rapprochements sont apparus entre les thèmes contraintes, graphes et fouille de données**. Le groupe a étudié l'apport des graphes et des contraintes pour améliorer la modélisation et la résolution de problèmes de fouille de données ; il a aussi établi des liens entre l'extraction de motifs et les hyper-graphes. Au-delà de ces rapprochements, notre ambition est une intégration de la fouille de données et des techniques d'optimisation dans les processus d'aide à la décision.

TT1 - Hybridation Fouille de Données et TAL.

RESPONSABLES : Thierry Charnois, Antoine Widlöcher.

PARTICIPANTS : T. Charnois, B. Crémilleux, F. Rioult, A. Widlöcher.

L'objectif de ce projet est de développer de nouvelles méthodes de fouille de données adaptées au matériau linguistique pour l'extraction de connaissances dans les textes. En effet, si les approches symboliques du traitement automatique des langues permettent des analyses fines du texte, elles s'appuient cependant sur des ressources coûteuses en temps de développement et peu génériques. De leur côté, les algorithmiques de fouille de données permettent de faire émerger des régularités sur des données volumineuses, mais ils ne sont actuellement pas adaptés aux données linguistiques. Ainsi, il apparaît de plus en plus nécessaire de repenser ces méthodes ou d'en élaborer des nouvelles, en fonction des spécificités de la donnée textuelle. Dans cette perspective, une hybridation profonde des méthodes de TAL et de fouille s'avère très prometteuse. Une telle hybridation vise d'une part à repenser la manière dont les processus de TAL peuvent tirer bénéfice de méthodes de fouilles spécifiquement adaptées au texte, et d'autre part à renforcer les processus de fouille par une prise en compte d'informations de nature linguistique visant notamment à sélectionner l'information la plus pertinente. Deux pistes seront simultanément exploitées : la fouille sous contraintes intégrant une modélisation linguistique et la fouille multi-échelles – en particulier la fouille de graphes – prenant en compte plusieurs niveaux de granularité du texte.

TT2 - Fouille de Graphes pour la Chimie.

RESPONSABLE : Bertrand Cuissart.

PARTICIPANTS : A. Bretto, B. Crémilleux, B. Cuissart, J.-J. Hébrard, F. Rioult, R. Bissell-Siders (postdoc), G. Poezevara (doctorant).

Ces travaux participent à la mise en place de la nouvelle réglementation européenne concernant la production, la commercialisation et l'utilisation des produits chimiques sur le territoire européen, la réglementation REACH. Plus particulièrement, ces travaux sont destinés à prédire l'impact sanitaire et environnemental de l'utilisation d'un composé chimique. Vu le nombre de composés chimiques à traiter, il est impossible d'appliquer des méthodes traditionnelles *in vivo* (tests répétés dans des laboratoires agréés). En conséquence, REACH permet d'utiliser des méthodes alternatives dont des méthodes informatiques (méthodes qualifiées de *in silico*). Les nouveaux outils développés par l'équipe vont être appliqués pour la prédiction de la toxicité aiguë, notamment sur les petits mammifères.

Par ailleurs, le monde de la chimie doit rapidement appliquer les principes de la "chimie verte". La "chimie verte" implique notamment de *substituer* les matières polluantes ou dangereuses par des matières ayant la même utilisation, mais ayant un meilleur impact sanitaire et environnemental. Le problème de la substitution des matières plastiques fera prochainement l'objet d'une collaboration avec des partenaires industriels. Dans ce cadre, nous nous intéresserons aux réactions chimiques comme objets ; sur un plan informatique, ce problème nécessitera de développer des techniques de fouille lorsque les données sont modélisées par des graphes dynamiques.

TT3 - Méthodes à base de Contraintes pour la Fouille de Données.

RESPONSABLE : Patrice Boizumault.

PARTICIPANTS : P. Boizumault, B. Crémilleux, A. Lallouet, S. Loudni, F. Rioult, J.-P. métivier (postdoc) et M. Khiari (doctorant).

Les méthodes actuelles de fouille sous contraintes estiment généralement la qualité d'un motif suivant les propriétés intrinsèques de celui-ci. Il s'agit d'une limite importante. La Programmation par Contraintes offre un cadre unique pour modéliser de façon déclarative et résoudre de nombreux problèmes combinatoires. Nous avons récemment proposé la notion de contraintes n-aires en fouille de données. Une telle contrainte porte sur plusieurs motifs et les relations liant les motifs entre eux confèrent alors un intérêt "global" aux motifs finalement sélectionnés. Il est clairement intéressant d'étudier les liens entre la découverte de motifs sous contraintes et les travaux menés en programmation par contraintes. Les objectifs sont multiples.

- Un premier axe est la définition de contraintes portant sur plusieurs motifs pour produire les modèles demandés par les utilisateurs (e.g., classifieurs, clustering, découverte de sous-groupes) et quelles méthodes de résolution mettre en œuvre. Pour cela, nous pensons que l'apport des CSPs quantifiés, qui permettent l'usage de quantificateurs universels, sera précieux. Afin de mieux comprendre la notion de contraintes en fouille, nous pensons utile de chercher à établir une typologie des contraintes (e.g., locale, n-aire, globale) et nous étudierons les liens entre celle-ci et les problèmes de satisfaction et d'optimisation afin de mettre en œuvre les méthodes appropriées de résolution.

- Un deuxième axe porte sur le savoir-faire des contraintes globales des CSPs afin de définir de nouveaux usages des contraintes pour la fouille. Par exemple, est-ce que la contrainte **regular** des CSPs est utile dans la fouille de séquences telles que les séquences de promoteurs de gènes, le texte, le web mining ?

- Un troisième axe concerne l'optimisation pour la fouille de données. Si les modèles utilisés par les méthodes de décision (par exemple la classification supervisée) peuvent utiliser des connaissances obtenues par la fouille de données, la mise au point de ces modèles, pour les rendre opérationnels, requiert des techniques d'optimisations multi-critères. Cependant, ces critères se focalisent sur la qualité du résultat, mais ne prennent pas en compte les contraintes spécifiques liées aux données, par exemple lorsque les classes sont déséquilibrées, ou liées à l'application du modèle, selon l'application visée. Nous souhaitons donc explorer des approches formalisant la conception d'un modèle de classification à base de règles d'association sous forme de problèmes de satisfaction ou d'optimisation sous contraintes.

- Enfin, un quatrième axe porte sur la conception de langages déclaratifs de haut niveau pour exprimer et représenter les requêtes de fouille de données manipulant données et motifs et portant sur des problèmes tels que l'énumération de motifs intéressants ou la notion de couverture.

TT4 - Décomposition de graphes et Contraintes.

RESPONSABLE : Samir Loudni.

PARTICIPANTS : P. Boizumault, J.-J. Hébrard, A.Lallouet, S. Loudni, X. Dupont, M. Fontaine (doctorants).

Malgré l'existence de problèmes réels fortement structurés, la structure du graphe sous-jacent à un réseau de contraintes est actuellement très peu prise en compte dans la résolution des problèmes de satisfaction et d'optimisation sous contraintes. Une méthode de décomposition arborescente décompose un graphe en composantes fortement connexes, appelées clusters, et les organise sous forme d'un graphe acyclique. Au niveau de granularité des cluster, la recherche de solution devient donc un problème polynômial.

Alors que ces décompositions n'ont été utilisées qu'au sein des méthodes de recherche complètes, nous proposons de les exploiter au sein des méthodes de recherche locale, notamment VNS (Variable Neighborhood Search), afin de guider l'heuristique de choix de voisinage. Deux principale directions de recherche seront considérées : i) exploiter la décomposition en clusters pour définir des structures de voisinage. Toutefois, l'ordre dans lequel les clusters seront considérés peut avoir une importance majeure. La situation du cluster dans le graphe peut nous renseigner sur l'influence de l'affectation de celui-ci sur ses voisins. D'autres stratégies tenant compte de la marge de progression du cluster restent à définir ; ii)

plutôt qu'exploiter la décomposition à priori, on cherche à guider la recherche locale en considérant des sous-parties décomposées du graphe initial.

2.2.2 Projets propres aux "Cœurs de métiers"

CMG - Graphes.

RESPONSABLE : Jean-Jacques Hébrard.

PARTICIPANTS : A. Bretto, B. Cuissart, J.-J. Hébrard, R. Bissell-Siders (post-doct), C. Jaulin et Y. Silvestre (doctorants).

L'étude concernant les G-graphes ayant été très fructueuse, nous envisageons de la poursuivre, en insistant sur les applications : poursuivre l'étude spectrale des G-graphes et le problème inverse, associer un groupe à un graphe. Dernièrement nous avons introduit un groupe à partir d'un graphe : un graphe donne naissance à un groupe associé avec un système de générateurs de telle sorte que deux graphes sont isomorphes si et seulement si les deux groupes sont isomorphes. Ces groupes sont très proches des groupes de Coxeter et leurs propriétés commencent à être dégagées. L'action de ce groupe sur l'espace vectoriel associé au graphe fournit des classes d'équivalences qui seront utilisées pour l'identification des graphes (isomorphes), notamment au sein des processus de fouille de graphes.

Au niveau de la conception algorithmique, nous allons poursuivre l'étude des appariements de graphes et des *contractions d'arêtes*. Cette dernière étude est principalement destinée à être appliquée au traitement des images numériques.

Une autre perspective porte sur les hypergraphes et la segmentation. L'introduction des hypergraphes en analyse d'images a donné des résultats conséquents. Cette modélisation est devenue un thème de recherche actif et une petite communauté internationale s'est créée autour de ce concept. Nous avons introduit une notion d'hypergraphe orienté permettant de modéliser les images satellites. Le but de notre recherche est de détecter des objets dans des images satellites grâce à de nouvelles techniques de classification basées sur les hypergraphes, ce qui est un enjeu important en analyse d'images haut niveau. Ce travail s'effectuera en collaboration avec TelecomParisTech et le CNES.

CMC - Contraintes.

RESPONSABLE : Samir Loudni.

PARTICIPANTS : P. Boizumault, A.Lallouet, S. Loudni, X. Dupont, M. Fontaine et M. Khiari (doctorants).

– *Relaxation contraintes globales*. L'objectif de ce travail est de définir de nouvelles fonctions de coûts globales ayant une sémantique pertinente et permettant de combiner les techniques de filtrage des WCSP avec celles des contraintes globales relaxées. Deux principale directions de recherche seront considérées : exploiter des cohérences locales plus sophistiquées telles que EDAC et VAC et étudier l'apport de la contrainte globale **regular** pour capturer des fonctions de coût d'arité quelconque.

– *Contraintes quantifiées*. En profitant de la plate-forme QeCode conçue par l'équipe, nous avons de multiples voies de recherche ouvertes : amélioration du branch&bound quantifié, développement d'heuristiques, parallélisation du solveur, jeu à n joueurs, etc. Tous ces problèmes comportent de véritables défis en terme de complexité spatiale et temporelle.

– *Contraintes sur les flux de données*. L'objectif de ce projet est de proposer un cadre pour la modélisation et la résolution de problèmes sur les flux infinis. Lié au model-checking et à la simulation, ce cadre nous permettra de nous intéresser à l'interaction et au couplage automate/résolution online pour modéliser des systèmes réactifs et interactifs.

CMF - Fouille de Données.

RESPONSABLE : Bruno Crémilleux.

PARTICIPANTS : B. Crémilleux, B. Cuissart, F. Rioult, R. Bissell-Siders (post-doct), L. Ben Othman, M. Khiari et G. Poezevara (doctorants).

Les méthodes actuelles de fouille sous contraintes estiment généralement la qualité d'un motif suivant les propriétés intrinsèques de celui-ci. Il s'agit d'une limite importante. Une perspective, amorcée par notre proposition de contraintes n-aires (cf. section 2.2.1), est de proposer un cadre général pour évaluer la plus-value d'un motif par rapport à d'autres motifs ainsi qu'une méthode de résolution. Pouvoir définir des contraintes portant sur plusieurs motifs, construire des motifs globaux offrent une vision synthétique des données à partir de motifs locaux afin de capturer leur richesse et leur complémentarité, restent des défis

majeurs en fouille de données. Une approche complémentaire concerne l'apprentissage de contraintes. En effet, la plupart des recherches en fouille de données sous contraintes ne se préoccupent pas de la question de l'origine des contraintes. Est-il possible d'aider un analyste dans son processus de requêtes en le guidant lors de la définition des contraintes relevant du domaine de son problème, voire par un apprentissage de contraintes appropriées ?

Un autre axe est la découverte de motifs qui prennent en compte les niveaux de granularité qui sont présents dans un document. Cette démarche s'inscrit dans une approche à plus long terme où il s'agit de dégager des dimensions structurelles sur les données comme par exemple leur séquentialité, leur granularité, leur structuration (e.g., XML, graphes) pour produire des méthodes de fouille reposant sur ces dimensions et non liées à la représentation "classique" (proche de la structure physique) des données. Enfin, une thématique naissante développe des applications de la fouille de données pour l'analyse stratégique dans le sport et le sport électronique. Le sport électronique regroupe les pratiques individuelles ou collectives liées au jeu vidéo en compétition. L'étude des traces de sport électronique dégage des perspectives encourageantes pour l'adaptation de ces méthodes d'analyse au sport réel.

2.3 Vie de l'équipe

Afin de faire vivre l'équipe et tant que sa taille le permet, il est prévu d'organiser une demi-journée hebdomadaire de discussion autour d'un thème lié à l'équipe. Il peut s'agir du suivi d'un étudiant en thèse ou en Master recherche, de la répétition d'une présentation orale de conférence, ou simplement d'un compte-rendu de lecture d'un article récent et intéressant.

Nous avons indiqué en introduction les deux séminaires hebdomadaires "historiques" du laboratoire susceptibles d'intéresser les membres de l'équipe : le séminaire Algorithmique et le séminaire I3. Toutefois aucun des deux ne porte spécifiquement sur les thématiques de CoDaG. L'équipe organisera un séminaire invité environ tous les mois en collaboration avec les deux séminaires existants, qui aura lieu pendant la demi-journée de discussion.

L'équipe possède actuellement un site collaboratif permettant de communiquer, de stocker et d'échanger des documents. Il est prévu de structurer ce site et d'encourager son utilisation.

Annexe-1 : Projets en cours

- **Projet ANR Bingo2** (Knowledge discovery For and By Inductive Queries in post-genomic applications) suite à l'appel à projets "Masse de données et connaissances" (janvier 2008 à décembre 2011). Partenaires : CGMC (CNRS UMR 5534, Lyon), GREYC (CNRS UMR 6072), LaHC (CNRS UMR 5516, Saint-Etienne) et LIRIS (CNRS UMR 5205, Lyon). <http://bingo2.greyc.fr/>. Montant total : 320.000 €. Montant pour le GREYC : 97.074 €. Coordinateur : B. Crémilleux (GREYC). PARTICIPANTS : R. BISSELL-SIDERS, P. BOIZUMAULT, T. CHARNOIS, B. CRÉMILLEUX, B. CUISART, M. KHIARI, G. LEJEUNE, J-L. MANGUIN, N. LUCAS, G. POEZEVARA
- B. CUISART est membre du **Projet ANR Innotox** (Validation of *in silico* and *in vitro* methodologies for the evaluation of Toxicity and Ecotoxicity of substances and preparations) suite à l'appel à projets « Chimie et Procédés pour un Développement Durable » (décembre 2007 - novembre 2010). Ce projet rassemble la société PCAS, les laboratoires CERMN (Université de Caen, UPRES EA 4258), Ecotoxicité Santé Environnementale (Université de Metz, UMR CNRS 7146) et associe le GREYC.
- **Projet Innotox2** (développement de méthodes alternatives de prédiction de la toxicité) suite à l'appel à projets "Emergence" du Conseil Régional de Basse-Normandie). Janvier 2009 à décembre 2011. Partenaires : CERMN (UPRES EA 4258), GREYC (UMR CNRS 6072), PE2M (UMR 100 IFREMER). Coordinateur : M-P. Halm (CERMN). Montant total : 230.200 €. Montant pour le GREYC : 86.768 €. Coordinateur pour le GREYC : B. Cuissart.

La thèse de Guillaume Poezevara (10/2008-10/2011) a pour objet de développer un algorithme de découverte de sous-graphes particuliers : les toxicophores. Dans le cadre du projet INNOTOX2, un travail postdoctoral a débuté en mars 2009 pour une durée de 24 mois, il a pour objet l'étude de deux descripteurs de graphes moléculaires dans un cadre de prédiction d'activité physico-chimique.

PARTICIPANTS : R. BISSELL-SIDERS, B. CRÉMILLEUX, B. CUISART, G. POEZEVARA

- **Projet ANR Asap** ("Apprentissage Statistique par une Architecture Profonde") suite à l'appel à projets "Domaines Emergents Programme Phare Mémoire" (janvier 2010 - décembre 2012). Partenaires : LITIS (EA 4108, Rouen), LIP 6 (CNRS UMR 7606, Paris), LIF (CNRS UMR 6166, Marseille), INRIA (équipe recherche TAO, Saclay), GREYC (UMR CNRS 6072, Caen). Coordinateur : Alain Rakotomamonjy (LITIS). Coordinateur pour le GREYC : C. Rosenberger. Montant total : 667.000 €. Montant pour le GREYC : 120.020 €. PARTICIPANTS : C. CHARRIER, B. CRÉMILLEUX, B. HEMERY, R. GIOT, C. ROSENBERGER

Annexe-2 : Sélection de 10 publications communes

Revue :

- 1 C. Hébert, A. Bretto, and B. Crémilleux. A Data Mining Formalization to Improve Hypergraph Minimal Transversal Computation. *Fundamenta Informaticae*, 80(4) :415–433, 2007. IOS Press.
- 2 M. Plantevit, T. Charnois, J. Kléma, C. Rigotti, and B. Crémilleux. Combining Sequence and Itemset Mining to Discover Named Entities in Biomedical Texts : A New Type of Pattern. *Int. J. of Data Mining, Modelling and Management*, 1(2) :119–148, 2009.
- 3 T. Charnois, M. Plantevit, C. Rigotti, and B. Crémilleux. Fouille de données séquentielles pour l'extraction d'information dans les textes. *Traitement Automatique des Langues*, 50(3) :30p, 2009.

Conférences Internationales :

- 1 R. Bissell Sidors, B. Cuissart, and B. Crémilleux. On the stimulation of patterns Definitions, calculation method and first usages. In M. Croitoru, S. Ferré, and D. Lukose, editors, *18th International Conference on Conceptual Structure (ICCS'10), Conceptual Structures : from information to intelligence*, volume 6208 of *Lecture Notes in Artificial Intelligence*, pages 56–69. 2010.
- 2 P. Cellier, T. Charnois, and M. Plantevit. Sequential Patterns to Discover and Characterise Biological Relations. In *11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'10)*, volume 6008 of *LNCS*, pages 537–548, Iasi, Romania, 2010.
- 3 P. Cellier, T. Charnois, M. Plantevit, and B. Crémilleux. Recursive Sequence Mining to Discover Named Entity Relations. In *9th International Symposium on Intelligent Data Analysis (IDA'10)*, volume 6065 of *LNCS*, pages 30–41, Tucson, Arizona, United States, 2010.
- 4 M. Khiari, P. Boizumault, and B. Crémilleux. Combining CSP and Constraint-based Mining for Pattern Discovery. In *International Conference on Computational Science and Its Applications (ICCSA'10)*, volume 6017 of *LNCS*, pages 432–447, Fukuoka, Japan, 2010.
- 5 M. Khiari, P. Boizumault, and B. Crémilleux. Constraint programming for mining n-ary patterns. In *16th International Conference on Principles and Practice of Constraint Programming (CP'10)*, volume xx of *LNCS*, St Andrews, Scotland, 2010. 15 pages, to appear.
- 6 S. Lozano, G. Poezevara, B. Cuissart, M-P. Halm Lemeille, E. Lescot Fontaine, A. Lepailleur, B. Bureau, and S. Rault. Assesment of chemical risk phrases in ecotoxicology : comparison of two methods. In *14th International Symposium on Toxicity Assessment (ISTA'09)*, Metz, France, 2009.
- 7 G. Poezevara, B. Cuissart, and B. Crémilleux. Discovering Emerging Graph Patterns from Chemicals. In *18th International Symposium on Methodologies for Intelligent Systems (ISMIS'09)*, volume 5522 of *Lecture Notes in Artificial Intelligence*, pages 45–55, Prague, Czech Republic, 2009.

3

Équipe Modèles, Agents, Décision

3.1	Équipe MAD	39
3.1.1	Composition de l'équipe	39
3.1.2	Perspectives scientifiques	40
3.1.2.1	Thèmes et projets communs	40
3.1.2.2	Thème 1 : Modèles	41
3.1.2.3	Thème 2 : Agents	41
3.1.2.4	Thème 3 : Décision	42
3.1.3	Collaborations avec d'autres équipes et disciplines	43
3.1.3.1	Projet ANR Défi Carotte	43
3.1.3.2	Projet Régional RASTARAD (2010–2013) et ANR CRITERE (2010–2013)	43
3.1.3.3	Collaborations avec l'équipe Automatique	44

Équipe MAD

Modèles, Agents, Décision

Nouveau responsable : Bruno Zanuttini

3.1 Équipe MAD

3.1.1 Composition de l'équipe

Sur la période 2006–2010 l'équipe comptait 8 permanents, tous producteurs. L'équipe sera renforcée à partir de la rentrée 2010 par Grégory Bonnet, recruté comme maître de conférences. Les autres membres permanents resteront dans l'équipe. Cependant, la délégation de Serge Stinckwich à l'Institut de Recherche pour le Développement au Vietnam est renouvelée pour 2010–2012.

Bruno Zanuttini deviendra le directeur de l'équipe en septembre 2012.

L'équipe restera structurée en trois thèmes : **Modèles (T1)**, **Agents (T2)**, **Décision (T3)**. Les permanents qui la composeront sont listés dans le tableau suivant.

Nom	Prénom	Grade	Thèmes	Date d'entrée
Bonnet	Grégory	MC	T2	2010
Bourdon	François	Professeur	T2	1997
Bouzid	Maroua	Professeur	T1, T2	2002
Jeanpierre	Laurent	MC	T3	2004
Mermet	Bruno	MC	T2	2006
Mouaddib	Abdel-Allah	Professeur	T1, T2, T3	2001
Simon	Gaële	MC	T2	2006
Stinckwich	Serge	MC	T2	1994
Zanuttini	Bruno	MC	T1, T3	2004

Risques L'équipe sera donc composée de 9 membres permanents, dont 5 ont une activité d'enseignement délocalisée : 3 à l'IUT de Caen (Campus III, à l'opposé de l'agglomération) et 2 au Havre. Serge Stinckwich (en poste à l'IUT) est en délégation au Vietnam jusqu'en 2012. Pour l'équipe, ceci implique que les enseignements en Licence et en Master 1 sont réduits, avec le risque d'une baisse d'attractivité auprès des étudiants (et une moindre connaissance des étudiants par les membres de l'équipe). En outre, la situation de Serge Stinckwich éloigne ses activités du reste de l'équipe. Nous comptons sur le recrutement de Grégory Bonnet pour compenser cet affaiblissement du thème « Agents ».

Opportunités

- **Renforcement de l'équipe** La première opportunité est le recrutement de Grégory Bonnet, sur les thématiques de la coordination multi-agents et de la sécurité dans les réseaux pair-à-pair. Cette dernière thématique s'inscrit parfaitement dans les préoccupations déjà établies du thème « Agents ». Quant aux problématiques de coordination, elles renforceront le lien entre les thèmes « Modèles » et « Décision » d'une part, et « Agents » d'autre part, notamment via l'utilisation de formalismes logiques (promesses entre agents par exemple). Cette thématique permettra à l'équipe d'aborder de façon plus globale la problématique d'agents qui raisonnent pour décider *ensemble*.
- **Trois projets ANR en 2010** Une autre opportunité importante pour l'équipe est l'obtention, en mai 2010, de trois projets ANR blancs : LARDONS (porté par B. Zanuttini), CRITERE et TUPLES. Ces trois projets portent sur des thématiques centrales pour l'équipe : décision en présence d'informations numériques et symboliques, connaissances temporelles et spatiales, et analyse de l'efficacité pratique des planificateurs.
- **Deux thèses CIFRE** Deux thèses CIFRE débiteront en 2010 ou 2011, l'une en collaboration avec EADS (coencadrée par Maroua Bouzid et Abdel-Allah Mouaddib), et l'autre avec l'ONERA et le CNES (coencadrée par Abdel-Allah Mouaddib et Éric Bensana — ONERA).

- **Nouvelles collaborations** Ces projets et thèses, par leurs problématiques (détaillées dans les perspectives scientifiques), seront l’occasion pour l’équipe de continuer à se concentrer sur ses thématiques centrales. Ils permettront aussi de renforcer des collaborations nationales déjà importantes (LIRMM, LAMSADE, EADS) et d’en débiter de nouvelles (CRIL, LIP6, CNRS-LIVE, INRETS-MA et GEOSYSCOM, ONERA), et fourniront une vitrine importante pour les activités de l’équipe.

3.1.2 Perspectives scientifiques

Nos perspectives scientifiques pour 2012–2015 sont en continuité de nos travaux actuels, avec de nouvelles voies à explorer sur nos problématiques centrales, un renforcement de l’axe « Agents », et un accent fort mis sur des projets et des thèmes communs. La structure en trois thèmes restera la même. Nous continuerons donc à *étudier et proposer des méthodes pour permettre à un groupe d’agents adaptatifs, temps réel et sous contraintes de ressources, plongé dans un environnement dynamique, de prendre des décisions rationnelles pour la réalisation d’une mission.*

Pour ce qui est de l’animation, nous poursuivrons le groupe de travail hebdomadaire, au cours duquel les membres de l’équipe présentent tour à tour leurs travaux en cours, leurs résultats, des problèmes ouverts, etc. L’aspect informel des séances, par opposition aux séminaires, facilite l’interaction et l’échange d’idées ou de travaux parfois très préliminaires. Il favorise également la participation active des étudiants de master 2. Ce groupe de travail sera notamment un lieu privilégié pour travailler sur les thèmes et projets communs décrits ci-après, et pour susciter des collaborations concrètes entre les membres de l’équipe.

3.1.2.1 Thèmes et projets communs

Interaction homme-robots L’équipe a travaillé récemment sur plusieurs projets autour de l’interaction homme-robot, qui prend de plus en plus d’importance dans la robotique de service. Ces travaux ont essentiellement été menés dans le thème « Décision », avec des modèles fondés sur différentes versions des processus décisionnels de Markov, appropriés aux situations de robots compagnons, assistants ou assistés. Il apparaît opportun d’intégrer des représentations symboliques (logiques), en particulier pour la communication avec l’homme. De plus, les aspects de coordination doivent être envisagés, que ce soit du fait de l’implication de plusieurs robots, ou en voyant l’homme comme un agent du système. Ce projet est ainsi transverse aux trois thèmes de l’équipe.

Problèmes de décision structurés L’équipe a une forte reconnaissance internationale dans les thématiques de la décision, en particulier sur les modèles markoviens. Elle mène actuellement plusieurs travaux sur des problèmes structurés, que ce soit du fait d’une représentation compacte (utilisant la logique) ou d’une structure dans le système multi-agents (SMA) qui doit décider collectivement. Ces problèmes concernent ainsi les trois thèmes de l’équipe, le thème « Modèles » pour les représentations compactes, le thème « Agents » pour la structure du SMA, et le thème « Décision » pour les algorithmes.

Application et valorisation des résultats Afin de fournir une vitrine aux activités de l’équipe dans son ensemble, et ainsi rendre identifiable son positionnement thématique fort et ses résultats, nous mènerons trois projets. Tout d’abord, nous continuerons notre effort d’intégration sur des robots, car cela permet de prendre en compte les contraintes réelles (capteurs imprécis par exemple) et fournit des démonstrations convaincantes, même pour des non-experts.

Nous projetons également de mutualiser et mettre à disposition de la communauté des implémentations de nos algorithmes ainsi que nos plateformes (robots, grille de calcul), pour fournir une vitrine de nos activités, et pour faciliter la réutilisation de nos résultats, pour la recherche fondamentale comme pour les applications. Enfin, nous envisageons de développer une application concrète et commune mettant ces résultats en valeur. Une idée dans cette direction est un système multi-agents d’agendas personnels, capables à la fois d’apprendre des modèles de leurs utilisateurs (préférences sur des horaires de réunion par exemple), et de prendre des décisions coordonnées pour organiser en leur nom des réunions, rendez-vous, etc.

Ces trois projets vont dans la même direction, celle d’une mutualisation et d’une rationalisation de l’implémentation d’algorithmes, et de leur intégration dans des applications, au prix d’un travail

d'ingénierie déjà fourni par une partie de l'équipe. Ils fourniront de plus des sujets de projets pour les étudiants, renforçant ainsi l'attractivité déjà avérée grâce aux applications sur les robots.

3.1.2.2 Thème 1 : Modèles

Gestion des connaissances temporelles et spatiales entre entités hétérogènes Nous continuerons nos travaux sur l'information spatiale et temporelle, en proposant tout d'abord des extensions des formalismes pour la représentation d'informations imprécises, nécessaire dans les systèmes composés d'entités hétérogènes, comme par exemple l'interaction homme-robot dans le cadre du projet ANR Amorces, ou un système de médiation entre experts humains dans le cadre du projet ANR CRITERE.

- *Gestion des informations spatiales imprécises* Dans le cadre de la thèse de Lamia Belouaer, nous avons proposé une représentation de l'information spatiale imprécise utilisant une ontologie dédiée. Cette représentation se fonde sur des relations topologiques spatiales floues comme « à côté de », « au nord de », des relations métriques floues comme « loin », « proche », et permet d'inférer de nouvelles informations par la composition de ces relations. Nous comptons poursuivre en développant des techniques de raisonnement sur de telles informations.
- *Raisonnement spatial et temporel* Ce projet concerne les méthodes d'extraction automatique d'informations spatiales et temporelles, en mettant l'accent sur la robustesse aux erreurs, contradictions, redondances et imprécisions. Dans le cadre de la thèse d'Arnaud Saval (CIFRE avec EADS), nous nous sommes intéressés à la découverte de relations dans des informations non structurées grâce à une modélisation temporelle, spatiale et sémantique. Dans la continuité, et dans le cadre d'une nouvelle thèse CIFRE avec EADS, nous explorerons les méthodes de structuration automatique de connaissances dans l'objectif de les enrichir par la prise en compte de l'imprécision. Cette thèse permettra une collaboration avec l'équipe CoDaG, puisque Thierry Charnois sera sollicité dans l'encadrement.
- *Gestion des informations temporelles et spatiales entre agents hétérogènes* Dans le cadre du projet ANR CRITERE, nous étudierons l'échange d'informations spatio-temporelles entre deux entités hétérogènes, raisonnant sur des représentations différentes. Nous considérerons les échanges entre un expert et une machine (IHM), ou entre plusieurs personnes. Dans ce dernier cas, notre objectif est d'étudier un système de type médiation permettant à un chercheur, un expert et un décideur d'échanger des informations via une plateforme. Celle-ci doit être capable de fournir l'information au bon degré de granularité en fonction du destinataire, ce qui requiert une représentation riche, hiérarchique, avec différents niveaux de granularité, du plus expert au moins expert.

Préférences Nous poursuivrons nos travaux sur les préférences, en particulier sur l'apprentissage. Les travaux commencés ont notamment mis en évidence des scénarios jouets d'interaction entre un agent (apprenant) et un utilisateur. Nous revisiterons ces scénarios jouets dans le contexte réel de l'interaction homme-robot, dans le cadre du projet commun à l'équipe. Un prototype d'une application basée sur ces scénarios, dans le cadre d'une vente aux enchères, a été réalisé lors d'un projet de M1 en 2010.

Une autre perspective importante est l'agrégation, dans un contexte multi-agent ou multi-utilisateur, où une relation de préférence du groupe doit être construite à partir des préférences individuelles, par exemple pour satisfaire au mieux un groupe d'utilisateurs humains ou pour permettre à des agents de coopérer. Les questions sémantiques concernant la maximisation du bien-être social, et les questions algorithmiques qui en découlent, n'ont presque pas été étudiées par la communauté. Un stage de M2 recherche a été effectué sur ce sujet en 2010.

3.1.2.3 Thème 2 : Agents

Agents mobiles et réseaux pair-à-pair Nous poursuivrons le développement de notre approche sécurisée et répartie du stockage de documents, dans la continuité des thèses d'Hugo Pommier, qui a exploré l'idée d'un stockage réparti de documents sur les réseaux pair-à-pair, en se basant sur des algorithmes de nuées de fragments mobiles de documents (*flocking*), et de Benoît Romito, qui a permis de porter les algorithmes sur une couche réseau pair-à-pair maintenable et ayant de bonnes propriétés

(utilisation de SCAMP, basé sur les graphes aléatoires et les protocoles épidémiques) et de les valider expérimentalement sur un réseau de plus de cent machines.

Nous envisageons deux directions de recherche. Sur le plan théorique, nous utiliserons des résultats dans le domaine des codes correcteurs pour affiner les algorithmes de récupération des documents, en cas de panne ou d'attaque dans le réseau, mais aussi en caractérisant les propriétés nécessaires à la nuée en fonction du niveau de sécurité recherché et des contraintes réseau imposées. Sur le plan expérimental, nous déploierons nos algorithmes sur des grilles de calcul de plusieurs milliers de machines.

Spécification et preuve de comportements d'agents Nous poursuivons nos travaux sur la spécification et la preuve de comportements d'agents via les arbres de décomposition de buts (GDT), dans deux grandes directions. Nous développerons tout d'abord des outils permettant de mettre en œuvre et d'automatiser la méthode, étape nécessaire à son utilisation à grande échelle. Ces outils devront nous permettre de décrire aisément des GDT, d'établir des connexions effectives avec des prouveurs disponibles, de les exécuter et de les animer, et enfin d'analyser les échecs de preuve. L'animation de spécifications, qui permet de valider une spécification indéterministe avant de la raffiner, a été peu étudiée dans le domaine des SMA, malgré son intérêt pour des systèmes complexes. Quant aux échecs de preuve, notre but est de fournir des méthodes pour que la cause de l'échec soit explicite pour le concepteur du système, même s'il n'est pas expert en preuve automatique. Nous nous appuyerons pour ce faire sur la spécificité des obligations de preuve des GDT.

Après cette phase préliminaire, nous utiliserons et adapterons notre méthode pour des domaines applicatifs des SMA où le besoin de validation se fait sentir : la composition de services Web, en lien avec la thèse de Mohamad El Falou (soutenue en 2010), les interactions *business-to-business* (B2B), en lien avec la thèse de Sébastien Picant, et les systèmes embarqués, en lien notamment avec les expérimentations de l'équipe sur des robots. Dans tous ces domaines, s'assurer de la correction d'une composition ou d'une interaction est un problème important, car les services ou systèmes sont développés indépendamment les uns des autres. Dans ce cadre, la compositionnalité de notre méthode de preuve est un atout important.

3.1.2.4 Thème 3 : Décision

Problèmes de décision factorisés Nous poursuivons le travail commencé dans la thèse de Boris Lesner sur les problèmes de décision factorisés, en l'occurrence des processus décisionnels de Markov représentés de façon compacte, en utilisant la logique. Ce sujet est ainsi à cheval sur les thèmes « Modèles » et « Décision » de l'équipe, et s'intègre dans le projet commun, plus global, sur les problèmes de décision structurés. Une direction générale, qui sera prise dans le cadre du projet ANR LARDONS, consistera à mettre l'accent sur les questions sémantiques et algorithmiques soulevées par la cohabitation d'informations numériques (utilités, probabilités de transition entre états...) et d'informations logiques (description des états comme des affectations attribut-valeur, description d'états impossibles, de normes...). Dans ce contexte, nous étudierons à la fois le calcul de décisions, dans un environnement connu, et l'apprentissage par renforcement, dans un environnement initialement inconnu.

Observabilité partielle et décentralisation Nous continuerons à développer la thématique des processus décisionnels de Markov partiellement observables et décentralisés (DEC-POMDP), qui constitue l'une des spécialités de l'équipe. Les modèles proposés et développés dans les thèses d'Aurélien Beynier, de Mathieu Boussard et de Jilles Dibangoye (OC-DEC-MDP, 2V-DEC-MDP) s'appuient sur des types particuliers d'interaction, exploités pour la résolution, avec de grands succès tels que l'algorithme PBIP. Nous explorerons d'autres types d'interaction, tels qu'une observabilité jointe totale, une communication gratuite ou au contraire très coûteuse, etc.

Ainsi, dans la thèse d'Arnaud Canu (bourse DGA), nous prenons le contrepied de l'hypothèse communément admise selon laquelle les agents interagissent ensemble et tout le temps. Cette hypothèse complexifie le problème, et n'est pas réaliste. Nous étudions ainsi la reformulation d'un DEC-POMDP en un ensemble de problèmes représentant indépendamment les phases avec et sans interaction. Nous comptons explorer différentes hypothèses sur la structure des interactions, dans le cadre du projet commun à l'équipe.

Nous exploitons également la structure du problème dans la thèse de Guillaume Lozenguez (cotutelle avec le LASMEA, projet ANR R_Discover). Nous proposons de résoudre des DEC-POMDP comme un ensemble de POMDP faiblement couplés, en nous fondant sur des décompositions de graphes, naturelles

pour la cartographie. Nous continuerons à développer cette approche, notamment en étudiant les décompositions de façon plus approfondie et en intégrant des modèles hiérarchiques pour la représentation de l'espace.

Décision pour l'interaction homme-robot Nous poursuivrons nos travaux sur la décision de robots interagissant avec l'homme, dans le cadre du projet commun à l'équipe. Le positionnement de l'équipe, via la modélisation de l'ensemble du processus de décision, en tenant compte du rôle des robots comme de l'homme, est originale sur ce sujet porteur et récent.

En nous basant sur le formalisme proposé dans la thèse d'Abir-Béatrice Karami, nous étudierons trois problèmes précis. Tout d'abord, la réalisation d'une mission par un ou plusieurs robots sous la surveillance d'un opérateur humain (qui peut les contrôler) nécessite la prise en compte de sa concentration et de sa réactivité (pour reprendre le contrôle par exemple), qui peuvent être inégales dans le temps. Nous étendrons les modèles déjà proposés au cadre général de plusieurs missions, robots, et opérateurs humains.

Le second problème est celui de la réalisation conjointe de tâches par un robot et un homme. Ici, la connaissance des intentions de l'homme est essentielle au robot, pour pouvoir l'aider sans le gêner. Dans le cas où aucune conversation ne prend place, nous avons proposé des techniques permettant au robot d'inférer ces intentions à partir de ses seules observations. Un important travail reste à faire pour les rendre opérationnelles, ce que nous aborderons en utilisant des modèles hiérarchiques et topologiques pour réduire la complexité de la décision.

Le dernier problème concerne l'aide à l'homme pour la réalisation de tâches. Dans le cadre de la robotique de service, l'apprentissage des préférences et des intentions de l'humain prend place sur un grand nombre de tâches et de missions différentes, au fil des jours. Nous développerons le modèle proposé par Lætitia Matignon dans son post-doctorat, qui comprend un module épistémique, engagé dans la conversation avec l'homme, et un module décisionnel spécialisé pour les tâches. Ces modules sont basés sur des POMDP, mais nous comptons intégrer au premier des représentations plus naturelles pour l'homme, notamment des représentations symboliques de ses préférences, qu'il est par ailleurs possible d'apprendre efficacement (cf. projet « préférences » p. 41).

3.1.3 Collaborations avec d'autres équipes et disciplines

3.1.3.1 Projet ANR Défi Carotte

Le défi Carotte vise à cartographier une zone intérieure d'environ 120 mètres carrés, en 30 minutes maximum. Au cours de son exploration, divers objets doivent être localisés. Le module de décision que nous avons développé permet au robot de décider des zones à explorer de manière autonome. Les difficultés majeures de ce projet reposent sur l'aspect dynamique de l'environnement, sur sa découverte progressive, et surtout sur la gestion temps réel de la décision.

Les perspectives d'évolution sont nombreuses. En premier lieu, nous envisageons l'extension du modèle de décision à plusieurs robots. Les deux avantages majeurs de cette approche sont naturellement la rapidité de couverture de la zone, et la redondance matérielle permettant de pallier une éventuelle défaillance. En contrepartie, cette approche multiplie l'incertitude liée à la position des robots, et nécessite une stratégie de communication. En second lieu, nous envisageons de multiples interactions entre le module de décision et le module de détection d'objets à l'aide d'une caméra. L'objectif est de permettre au robot de se positionner pour prendre une photographie précise de son environnement. Une troisième amélioration anticipée repose sur la fusion de la localisation obtenue par le biais des capteurs à balayage LASER et de celle obtenue par la reconstruction 3D générée par les caméras panoramiques étudiées par l'équipe de Sophia-Antipolis. Nous projetons à ce titre une collaboration avec l'équipe Image (Frédéric Jurie en particulier).

3.1.3.2 Projet Régional RASTARAD (2010–2013) et ANR CRITERE (2010–2013)

Maroua Bouzid développe une recherche pluridisciplinaire entre géographes et informaticiens pour enrichir les SIG, afin de prendre en compte l'imprécision des informations spatio-temporelles. Cette collaboration consiste à enrichir les SIG, tout d'abord par une information spatiale floue, en se fondant sur des fonctions floues pour définir des relations spatiales topologiques ou de distance entre objets spatiaux. Pour cela, une ontologie, appelée *SpaceOntology*, est proposée dans la thèse de Lamia Belouaer. Cette

ontologie permet l'inférence de nouvelles informations, et donc l'alimentation du SIG. L'autre volet de cette collaboration consiste à enrichir les SIG par des informations concernant les risques, afin de permettre la définition de la vulnérabilité d'une zone, ou de son degré d'exposition à un ou plusieurs risques. La complémentarité entre la connaissance du risque (par les géographes) et la formalisation des informations permet d'aborder ce problème. Ces deux points sont étudiés dans les projets RASTARAD et ANR CRITERE (en collaboration avec GEOSYSCOM/UCBN et INRETS).

3.1.3.3 Collaborations avec l'équipe Automatique

Autour de la problématique du contrôle, les équipes MAD et Automatique du GREYC sont clairement complémentaires : décision et contrôle au niveau logiciel pour l'équipe MAD, en particulier en utilisant les modèles Markoviens, et contrôle des couches basses pour l'équipe Automatique. Nous envisageons dès la rentrée 2010 des groupes de travail qui nous permettront de lancer des projets communs, en particulier autour du contrôle des déplacements d'un robot (planification haut niveau pour relier des points bien identifiés, et contrôle des couches basses pour les relier effectivement). Ces projets pourront se concrétiser dans le cadre du projet Robots_MALINS, et nous envisageons également de proposer des stages de master à des étudiants en automatique autour du contrôle des robots Koala dont nous disposons.

4

Équipe Document numérique, langue, usages

4.0.4	Composition probable au 1er janvier 2012	46
4.0.5	Présentation	46
4.0.6	Auto-évaluation	47
4.0.6.1	Opportunités	47
4.0.6.2	Risques	48
4.0.7	Perspectives scientifiques	48
4.0.7.1	Thème 1 : Usages, interactions, sémiotique	48
4.0.7.2	Thème 2 : Sémantique, rhétorique, discours	49
4.0.7.3	Thème 3 : Multilinguisme, traduction, algorithmique du texte et méthodes différentielles	50
4.0.8	Organisation	51

Équipe DLU – Document numérique, Langue, Usages

2012-2015

4.0.4 Composition probable au 1er janvier 2012

Responsable : Nadine LUCAS (CR1, HDR).

En plus d'1 ou 2 PR recrutés en 2011, l'équipe comprendra en 2012 un effectif de 14 permanents, dont 3 pour moitié de leur activité, soit un équivalent de 12,5 permanents dont 11,5 producteurs : 9,5 MC, 1 CR1 CNRS, 1 IR et 1 PRAG. Un autre PR de l'équipe est en détachement depuis 2010 dans une université japonaise.

Prénom Nom	Fonction	Thèmes	HDR	Commentaire
Pierre Beust	MC UCBN	T1 T2		
Valérie Cauchard	MC UFR Sciences	T1		
Thierry Charnois	MC IUT Caen	T1 T2	prévue 2011	50% CODAG
Antoine Doucet	MC IUT Caen	T1 T2	prévue 2011	
Patrice Enjalbert	PR émérite	T1 T2		(3)
Stéphane Ferrari	MC UCBN	T2	prévue 2010	
Hervé Le Crosnier	MC UCBN	T1	oui	
Jean-Marc Lecarpentier	PRAG UFR Sciences	T1		doctorat prévu en 2011
Yves Lepage	PR UFR Sciences			en détachement depuis 2010
Nadine Lucas	CR1 CNRS	T2 T3	oui	
Jacques Madelaine	MC UFR Sciences	T1		50% AMACC
Jean-Luc Manguin	IR CNRS	T1 T2 T3		
Yann Mathet	MC UCBN	T1 T2		
Serge Mauger	MC IUT Caen	T1 T2		
Fabrice Maurel	MC UFR Sciences	T1 T2		
Jacques Vergne	PR émérite	T2 T3		(3)
Antoine Widlöcher	MC UCBN	T2		50% CODAG
X	PR UFR Sciences			sur poste Vergne ⁽¹⁾
Y	PR UFR Sciences			sur poste Enjalbert ⁽²⁾

⁽¹⁾ : Le poste est déclaré cette année sous le profil « Document numérique » et sera pourvu en septembre 2011 sous réserve d'acceptation de la demande par l'Université.

⁽²⁾ : Le poste est déclaré cette année sous un profil double « Traitement automatique des langues » (pour l'équipe DLU) ou « Protection de l'Information » (pour le Groupe PICC) et sera pourvu en septembre 2011 sous réserve d'acceptation de la demande par l'Université.

⁽³⁾ : Sous réserve d'acceptation de la demande par l'Université.

4.0.5 Présentation

L'équipe DLU *Document numérique, Langue, Usages*, provient de la réunion d'une partie de l'équipe ISLanD et d'une partie de l'équipe DoDoLa. Cette dernière équipe s'est divisée en deux entités, l'une autour de la fouille de données rejoignant CODAG, l'autre autour du TAL rejoignant DLU. L'équipe DLU regroupe l'ensemble des compétences du GREYC relatives au document textuel, et intègre des travaux sur les modalités texte-image. Le document numérique qu'elle étudie est à la fois multimodal, composite, multilingue, dynamique, perçu en interaction, dans ses usages, etc. L'un des objectifs de la nouvelle équipe sera de pérenniser les réalisations issues des deux équipes d'origine et de créer les conditions d'une réelle collaboration (cf infra *risques*).

L'ambition de cette équipe est de développer des coopérations autour des divers aspects du document numérique : génération, annotation et analyse de collections annotées ou non, usages. La plate-forme Document numérique et fouille du GREYC doit rendre visible le travail sur la fouille de textes et les collaborations avec l'équipe Image seront favorisées au sein du département Information et Reconnaissance des Formes.

Au niveau local, l'équipe devrait affirmer son rôle au sein du pôle Document numérique de la MRSH de Caen, rassemblant SHS et STIC, en partenariat avec l'équipe Image du GREYC notamment. L'équipe devrait s'investir dans le pôle TES autour des thématiques de dématérialisation. Des collaborations sur ce thème se dessinent en interne avec l'équipe Monétique et Biométrie. Au niveau régional, la Fédération NormaSTIC avec la Haute-Normandie devrait développer les projets applicatifs dans l'axe Langue et Interaction, un rapprochement initié au sein du Groupe Nu avec le LITIS de Rouen, entre autres. Les liens avec les entreprises locales sont déjà forts avec notamment Noopsis, start-up issue de Dodola, et se renforceront avec la start-up ?Meo! issue d'ISLAND.

Au niveau national, l'équipe est active dans le projet du Grand Emprunt Numérique Normandie (porté par la MRSH de Caen) dont un des aspects est la dématérialisation de documents avec une plateforme de numérisation et un *data center* à Caen. Elle a l'ambition de construire des projets labellisés dans la continuité de ses coopérations avec, à Toulouse, le CLE-ERSS et l'IRIT, à Paris, le LaTTiCe (UMR 8094), le LaLIC (Paris 4), à Orsay, le LIMSI, et à Lyon, l'ICAR (ENS Lyon).

Sur le plan international, elle cherchera à développer les questionnements sur la gestion de l'échelle, à animer des concours internationaux comme INEX *Book Search*, à renforcer une activité multilatérale avec des pays européens, suite au PICS avec la Finlande MultiPULS (2009-2011). Les relations avec le Japon où un PR (Yves Lepage) se trouve en détachement seront aussi favorisées. Les plate-formes existantes (LinguaStream, Glozz, Wims) seront maintenues ainsi que les outils du GREYC issus de plate-formes partagées (Calico). Une réflexion est à mener sur la compatibilité des outils produits avec ceux existant au niveau international (GATE, UIMA...).

Les travaux de l'équipe s'articulent autour de trois thèmes :

- Thème 1 : Usages, interactions, sémiotique – resp. Fabrice MAUREL
- Thème 2 : Sémantique, rhétorique, discours – resp. Stéphane FERRARI
- Thème 3 : Multilinguisme, traduction, algorithmique du texte et méthodes différentielles – resp. Antoine DOUCET

Les applications visées relèvent de l'ingénierie documentaire (accès au document et à son contenu, extraction et synthèse de l'information, aide à la lecture, à la navigation inter & intra documentaire) et de l'ingénierie des langues (aide à la traduction, résumé automatique, interfaces langagières, approches multidisciplinaires et outillage des Sciences Humaines et Sociales). Ces applications seront systématiquement envisagées en collaboration entre les trois thèmes de l'équipe.

4.0.6 Auto-évaluation

4.0.6.1 Opportunités

Le rapprochement des deux équipes Dodola et ISLAND a été recommandé par les évaluateurs. Il porte le nombre de membres permanents à 14. L'ancienne équipe ISLAND a été jusque là active et moteur d'innovation, ses productions touchent des utilisateurs au niveau international. L'ancienne équipe Dodola a plusieurs projets ANR et des utilisateurs au niveau national. Les atouts de la nouvelle équipe sont de rassembler des compétences fortes concernant le document numérique, depuis la génération jusqu'aux usages en passant par l'analyse. Elle a également une pratique de l'interdisciplinarité et des collaborations STIC-SHS, en particulier à travers le pôle Modesco de l'université de Caen. Les membres ont une bonne pratique de la publication jusqu'au niveau international et un seul est non produisant. La culture des équipes précédentes est néanmoins différente. Si les différences sont reconnues et acceptées, il sera possible de tirer parti de la complémentarité ou de l'alternance des approches, en particulier :

- sur le plan des méthodes, en confrontant les méthodes différentielles endogènes (sans ressources en mémoire), et les méthodes avec ressources, par exemple des approches sur corpus thématiques, appuyées par des grammaires locales, des ressources spécialisées ;
- sur le plan des grains d'analyse, avec des approches du texte entier, macro-syntaxiques, rhétoriques et stylistiques *versus* des approches phrastiques, ou au niveau du paragraphe, qui s'intitulent aussi rhétoriques et stylistiques.

4.0.6.2 Risques

Cette équipe est en formation : elle « doit faire ses preuves » dans un environnement peu favorable de compétition exacerbée. La reconnaissance et le respect des spécificités des approches des membres demandera beaucoup d'efforts à tous les niveaux. Des conflits d'intérêt sont possibles entre les entreprises issues des anciennes équipes. Un audit sera demandé sur les ressources logicielles relevant de la nouvelle équipe pour prévenir autant que possible les différends de propriété intellectuelle.

La place de l'équipe dans les projets collectifs devra être défendue avec vigilance et souci d'équilibre, au sein du département Information et Reconnaissance des Formes et au sein du pôle Document numérique. Les projets les plus hardis, potentiellement les plus porteurs à long terme, sont aussi ceux qui sont les plus difficiles à défendre, la concurrence étant vive sur les sujets sensibles au niveau international. Or, les instances locales et nationales exigent la caution internationale avant d'accorder la leur, cercle vicieux s'il en est. Il favorise le pillage des idées par les laboratoires étrangers et la fuite des cerveaux. Les collaborations actives sur la fouille textuelle, qui intéressent plusieurs membres dont 2 à mi-temps avec CODAG, doivent trouver un support en dehors du département, dans les projets liés à la plateforme Document numérique.

La période de transition avant le début officiel du quadriennal est délicate pour l'équipe DLU, car au moins un tiers de l'effectif de juin 2010 sera renouvelé d'ici janvier 2012. De nombreux départs prévus (retraites de deux Pr en 2011, mise à disposition d'un CR qui fonde une entreprise) affaiblissent l'équipe, ce qui s'ajoute à l'absence du Pr LEPAGE (en détachement au Japon depuis 2010). Ceci est à modérer cependant par des HDR en cours ou prévues. De nouveaux recrutements devraient aussi intervenir prochainement.

4.0.7 Perspectives scientifiques

4.0.7.1 Thème 1 : Usages, interactions, sémiotique

Ce thème a pour objectif de regrouper et d'organiser les activités de l'équipe qui abordent le document numérique sous l'angle d'un objet à la fois composite, interactif et sociétal.

Le lien qui réunit les différents intervenants de ce thème pourrait se définir par l'intérêt porté à l'objet « document numérique », non pas en tant que regroupement d'informations langagières numérisées, mais en tant qu'objet sémiotique en constante évolution, composé d'éléments multiples et multiformes et avec lesquels les utilisateurs interagissent.

Par nature, la plupart des aspects informatiques abordés dans ce thème le sont de manière pluridisciplinaire en faisant appel à des connaissances linguistiques, psychologiques, géographiques, philosophiques ou encore issues des sciences de l'éducation.

Dans ce cadre, le premier enjeu (I) est d'étudier la manière avec laquelle un document numérique peut se composer en combinant différents médiums, différentes modalités d'accès ou de production, différentes présentations des informations, etc. Ces aspects sont abordés à travers (1) l'élaboration d'atlas géographiques, (2) l'implémentation d'un nouveau modèle de document numérique issu des méthodes et réflexions des milieux des bibliothèques et des musées et (3) l'étude des relations entre le texte et l'image.

Un second enjeu (II) concerne l'interaction entre les documents numériques et les acteurs impliqués à la fois dans leur élaboration et dans leur utilisation. Nous traiterons de ces questions à travers (1) un projet d'environnement numérique de travail (ENT) éactif, (2) l'élaboration de services multimodaux à partir de la notion d'interface plastique et (3) l'étude de la « transmodalisation » et la « multimodalisation » de l'information.

Enfin, le troisième enjeu (III) que nous traiterons concerne la dimension sociale et sociétale du document numérique. Nous aborderons cet aspect par (1) la mise en oeuvre et l'expérimentation, dans le cadre de l'échange de documents entre entreprises, de protocoles répartis pour le référencement et la découverte, (2) l'anticipation et l'accompagnement des usages en analysant les données (traces) produites consécutivement à l'activité des individus et qui reflètent le contexte de cette activité et (3) la définition des nouvelles notions de « culture numérique » et de « biens communs » abordée d'un point de vue médiologique. La structuration de ces enjeux est résumée ci-dessous.

- I. Le document numérique composite
 - (1) Élaboration d'atlas géographique
 - (2) Représentation des documents numériques composites et multilingues
 - (3) Relations texte-image

- II. Le document numérique interactif
 - (1) Vers un ENT éactif
 - (2) Élaboration de services multimodaux
 - (3) Transmodalité et multimodalité
- III. Le document numérique dans la société
 - (1) Échange de documents : référencement et découverte – *Sydonie*
 - (2) Études de traces *Calico*
 - (3) Culture numérique et biens communs – *Calico*

On décrit ci-dessous deux actions en cours que l'on veut développer lors du prochain quadriennal.

Framework Sydonie Dans le cadre du Framework Sydonie (Système pour les documents numériques, l'internet et l'édition), nous travaillons avec les bibliothèques à la représentation des documents numériques composites et multilingues.

INEX BookSearch, collections d'ouvrages numérisés Depuis sa création en 2007, nous participons à l'organisation du forum international d'évaluation de recherche d'information (RI) dans des collections d'ouvrages numérisés *Book Search*. Il s'agit du seul effort d'évaluation académique des recherches à l'échelle du livre, utilisant une collection de plus 50 000 ouvrages numérisés libres de droit. Son objectif est entre autres de valider les principes de la RI structurée dans des bibliothèques numériques. Le choix du grain de la réponse (livre, chapitre, partie, paragraphe, passage?) est l'une des questions abordées. Dans un contexte de numérisation massive, cette initiative a très vite pris de l'essor : le nombre d'institutions participantes est passé de 27 en 2007, à 84 en 2009. Les participants sont invités à présenter leurs approches dans des actes publiés par Springer (*Lecture Notes in Computer Science*).

Pour traduire les documents numérisés (images) en documents structurés (texte balisé), nous avons également lancé en 2008 la tâche *Structure Extraction* qui s'adresse à la communauté de l'analyse documentaire, qui intervient en amont de la communauté RI. Compte tenu de l'essor des processus de numérisation, nous espérons dans les années à venir développer les collaborations avec les bibliothèques nationales et faire déboucher ce travail sur des projets significatifs.

4.0.7.2 Thème 2 : Sémantique, rhétorique, discours

Le thème « sémantique, rhétorique, discours » a pour objectif d'organiser les activités de l'équipe relatives au traitement automatique du contenu textuel des documents. Les travaux qui y sont menés concernent l'analyse du sens, l'aide à l'interprétation, ou encore l'articulation argumentative.

Démarche, méthodes Les travaux passés ont permis de mettre en place des collaborations fortes avec des équipes en Sciences Humaines et Sociales, notamment en linguistique. Nous comptons poursuivre cette démarche, directement liée à notre méthode de travail : il s'agit en effet de tenir compte des connaissances qu'apportent les SHS sur les objets d'étude que nous avons en commun, afin de proposer des modèles formels qui s'en inspirent. De manière générale, les travaux suivent une démarche en trois temps : (i) analyse de l'objet d'étude – *p. ex. travail sur corpus* ; (ii) proposition, élaboration d'un modèle informatique – *inspiré d'un modèle linguistique* ; (iii) mise en œuvre et confrontation à des données textuelles – *pour validation et retour sur le modèle ou pour évaluation*.

Les évaluations quantitatives sont souvent coûteuses, nécessitant à la fois une réflexion sur les protocoles, les mesures, et une mise en place de corpus de référence souvent inexistantes pour les thématiques abordées. Nous développerons plus avant la réflexion sur ce point, central à de nombreux travaux du thème. Nous accentuerons les efforts de développement et de valorisation des plates-formes logicielles (LINGUASTREAM, GLOZZ, CALICO...).

L'exploitation de ressources sémantiques pose quant à elle un problème de fond. Une réflexion sera menée dont l'objectif est de mettre en exergue et d'expérimenter des contextes d'utilisation de ressources termino-ontologiques dans une approche individu centrée, c'est-à-dire des ressources sémantiques non exhaustives mais très ciblées autour des activités d'un sujet, et de ce fait légères (en terme d'espace, de traitement et de maintenance) et facilement partageables. Le développement de ressources et d'outils mettra à contribution les méthodes de fouille de données en interaction avec l'équipe CODAG.

Structuration textuelle, thématique ou argumentative L'objectif est de reconnaître un mode d'organisation des textes afin de permettre par exemple une aide à la lecture, la recherche d'information ou l'extraction d'information. Plutôt adaptés aux documents longs (ouvrages, articles scientifiques), ces travaux trouvent aussi leur application dans des corpus de nouvelles journalistiques ou des collections de documents techniques. Initiée dans les projets **GeoSem** (2002-2006) puis l'**ANR Annodis** (2007-2010), l'analyse de la structuration du discours vise à mettre en lumière l'organisation des textes.

En particulier, une réflexion est menée sur la notion de *zoning* rhétorique. Elle vise à étudier l'articulation entre différents modes d'organisation du texte, notamment sur les plans argumentatif et thématique. Ce travail novateur fait actuellement l'objet d'une collaboration étroite avec le laboratoire LaTTiCe (ENS Ulm, Paris), qui devrait s'étendre à d'autres équipes au niveau international. Plusieurs cadres applicatifs seront testés pour valider nos modèles. Citons en particulier celui offert par le projet **ANR ProDescartes** (2009-2013), « Corpus Descartes. Projet d'édition en ligne des œuvres et de la correspondance de Descartes ». Rassemblant des philosophes de Caen, de Paris 4 et de Lecce (Italie), et des informaticiens du GREYC, avec l'appui des Presses Universitaires de Caen et de la société Noopsis, ce projet éditorial vise à établir une édition scientifique, en ligne, des écrits de Descartes. Cette édition tirera bénéfice de l'application au français classique et au latin, de méthodes d'exploration et d'analyse de corpus issues des modèles de structuration du discours élaborés aux GREYC.

Discours évaluatif, analyse d'opinion Une autre dimension de l'analyse du discours concerne son caractère objectif ou subjectif, qui doit être pris en considération dans les applications d'analyse de contenu. En particulier, la veille d'image ou le suivi d'opinion se concentrent sur le discours subjectif et évaluatif, pour analyser ou exploiter les jugements, appréciations et affects exprimés dans les textes. Si l'analyse informatisée des opinions répond à des enjeux économiques et sociaux importants, les travaux actuels se limitent souvent à la notion de polarité (positive ou négative), quand les phénomènes visés sont en réalité bien plus complexes. Notre objectif principal est de chercher à identifier et agréger des segments textuels porteurs d'opinions, de les classer par type (affect, jugement, appréciation, . . .) et de les caractériser finement : intensité, prototypicalité, cible, source, engagement de l'auteur, etc.

Les travaux menés sur cette thématique feront suite au projet ANR OntOpiTex (« Modèles linguistiques et ontologies. Extraction informatique et caractérisation d'opinions et de jugements d'évaluation dans les textes », 2009-2012, 3 ans), qui regroupe linguistes, informaticiens et industriels sur cette problématique : LaLIC (Université Paris IV), GREYC et CRISCO (Caen), sociétés Noopsis (Caen) et TecKnowMetrix (Grenoble). Au-delà de ce projet, nous poursuivrons l'étude de cette problématique en la croisant notamment avec les travaux précédents sur les tropes et les figures de rhétorique, dont le rôle dans le discours évaluatif paraît central.

Temporalité C'est un sujet d'étude ancien du GREYC, structuré par une collaboration avec L. Gosselin (Linguistique, Rouen). Dans la prochaine période, nous pensons l'articuler autour des trois axes suivants : (I) *anaphore événementielle*. Repérage automatique des éléments anaphoriques dans le texte, et constitution des chaînes référentielles afférentes ; (II) *calcul automatique des valeurs aspectuo-temporelle des procès d'un texte* ; (III) poursuite du travail sur l'*itération temporelle*, et étude de l'*anaphore événementielle* dans ce cadre. Nous travaillerons en collaboration avec l'entreprise Noopsis sur les deux premiers axes dans le cadre d'applications en recherche et extraction d'informations.

4.0.7.3 Thème 3 : Multilinguisme, traduction, algorithmique du texte et méthodes différentielles

Analyse différentielle de discours Quoique l'objet d'étude soit le même que celui du thème 2, les méthodes différentielles représentent une rupture de paradigme. Les valeurs affectées ne sont pas absolues mais relatives. Les techniques calculatoires visent des procédures algorithmiquement efficaces sans ressources préalables. Les méthodes différentielles appartiennent aux approches « alternatives » dont font partie les méthodes dites *inductives*, *adaptatives* ou *systémiques* et elles s'inscrivent ainsi dans une frange émergente et fortement concurrentielle au plan international. Cette approche annule l'opposition entre méthodes symboliques et statistiques.

La réflexion a conduit à séparer clairement les principes (discursifs et rhétoriques) et les paramètres (le genre, la langue, le style individuel). Les applications déjà abordées sont la veille et l'extraction d'information massivement multilingues. La structuration des textes en terme de contenu informationnel

est aussi utile pour l'annotation de vastes collections de documents plein texte, de façon économe, pour la fouille de données textuelles (ACI Bingo2).

Une collaboration avec l'Université d'Helsinki dans le cadre du projet MultiPULs (2009-2011) sur la veille épidémiologique en 12 langues, se déroule en collaboration avec le Pr Yangarber (plate-forme européenne MedISys, projet PULS). À plus long terme, il faut assurer le lien entre les partenaires européens du projet MedISys en 45 langues et assurer la veille sur l'évolution des concepts. La résolution de l'anaphore pour le repérage d'évènements fait partie des prolongements naturels de ces travaux, ainsi que la production de synthèses (résumé multisources).

Dans les prochaines années, nous entendons valoriser nos méthodes, soit par des publications, soit par le dépôt de brevets. Nous entendons poursuivre la recherche de solutions fondées sur le repérage de structure au sens linguistique du terme. Du point de vue algorithmique, elles sont rebaptisées *méthodes géométriques de maillage du texte*, pour éviter les amalgames.

Traduction, alignement, segmentation La continuité des approches innovantes de l'ancienne équipe ISLanD et le suivi des contrats en cours sont une priorité, notamment le projet SAMAR sur la traduction de l'arabe jusqu'en 2012 (labellisé par le pôle de compétitivité Cap Digital, en collaboration avec le LIMSI). À plus long terme, l'orientation dépendra des recrutements futurs.

En traduction, les approches développées avec succès ont montré que les opérations de segmentation, d'alignement et de traduction sont seulement différents aspects d'un même problème et que tendre à la résolution conjointe de ces trois aspects est pertinent du point de vue de la recherche, et rentable en termes d'économie de moyens (en ressources linguistiques) et d'effort (c'est-à-dire de techniques de calcul à mettre en œuvre). Du point de vue méthodologique, il s'agit, dans le cadre d'une approche par l'exemple ou fondée sur les données, de proposer des techniques très économes en ressources et utilisant des résultats d'observation sur les langues.

En segmentation et alignement, l'objectif est d'optimiser le choix du grain permettant le meilleur alignement, par exemple sur les vastes collections européennes de textes, dans la lignée des travaux en cours avec l'entreprise Pertimm. Les alignements à gros grain, au grain article par exemple, sont une innovation. Une autre direction de recherche est l'alignement au meilleur coût sur une paire ou un petit nombre de langues, en exploitant les similarités observées. Les recherches futures s'appuieront sur les logiciels *Anymalign* et *Homomorphism* mis à disposition de la communauté. L'établissement de liens internationaux sera favorisé, notamment avec le Japon.

Algorithmique du texte et fouille de données textuelles Les approches endogènes, sans ressources lexicales en mémoire, sont indispensables pour des langues peu dotées (en ressources électroniques) et sont économes pour la recherche d'information dans des collections volumineuses. Elles produisent des descriptions mémorisables (production de lexique aligné, de mots-clés ou de lexies, de descripteurs de collections). De nombreuses tâches d'analyse de textes sont en relation avec la fouille de textes ou relèvent du diagnostic de texte ou de collections. Parmi les logiciels réalisés, l'extraction des séquences fréquentes maximales et les méthodes d'exploitation liées servent en recherche documentaire. Les recherches futures concerneront la définition d'algorithmes rapides pour travailler sur des collections de textes. Ces recherches supposent la définition de mesures de similarité nouvelles.

4.0.8 Organisation

L'équipe a un responsable élu ayant un rôle d'animation, d'application de la politique scientifique de l'équipe et de représentation au comité scientifique du laboratoire. Une réunion d'équipe aura lieu en moyenne toutes les 2 semaines. À l'occasion de ces réunions, une présentation de 20 minutes sera faite sur l'avancement d'une thèse, d'un projet ou d'une préparation à une conférence. La politique scientifique de l'équipe, les aspects budgétaires et les informations sur la vie du laboratoire y seront également discutés.

Équipe Monétique & biométrie

5.1	Equipe Monétique & Biométrie	53
5.1.1	Composition	53
5.1.1.1	Membres permanents	53
5.1.1.2	Membres non permanents	53
5.1.2	Auto-évaluation	54
5.1.2.1	Points forts	54
5.1.2.2	Points faibles	54
5.1.2.3	Opportunités	55
5.1.2.4	Risques	55
5.1.3	Perspectives scientifiques	55
5.1.3.1	Thème 1 : Monétique	56
5.1.3.2	Thème 2 : Informatique de confiance	56
5.1.3.3	Thème 3 : Biométrie	56
5.1.3.4	Aspects méthodologiques	57

5.1 Equipe Monétique & Biométrie

La création de l'équipe Monétique & Biométrie fait suite au groupe du même nom sur le quadriennal précédent. Cette évolution fait suite au recrutement des trois membres permanents au sein du GREYC sur ces thématiques et à sa montée en puissance sur la partie valorisation académique (la valorisation industrielle étant déjà un de ses points forts). Cette équipe se caractérise par sa très grande interaction avec le milieu industriel et le grand nombre de contractuels en son sein (dont le nombre est lié à la participation à de nombreux projets collaboratifs). Les thèmes de recherche (déjà présents au sein de ce groupe) contribuent à la sécurisation de transactions électroniques en s'attaquant à différents verrous scientifiques en collaboration avec des industriels et laboratoires de recherche. La dimension usage de ces nouvelles technologies est prise en compte avec la coopération active avec des chercheurs en psychologie cognitive.

5.1.1 Composition

5.1.1.1 Membres permanents

Prénom et nom	Fonction	Thèmes			Date d'entrée
		T1	T2	T3	
Patrick Lacharme	MC ENSICAEN		X		2010
Marc Pasquet	PR ENSICAEN	X			2009
Christophe Rosenberger	PR ENSICAEN			X	2007

TABLE 5.1 – Liste des membres permanents

5.1.1.2 Membres non permanents

Prénom et nom	Financement	Thèmes			Date d'entrée
		T1	T2	T3	
Refka Abdellaoui	CIFRE Limonetik	X	X		09/2008
Vincent Alimi	CIFRE Inside Contactless	X	X		11/2009
Rima Belguechi	Co-tutelle Algérie			X	01/2010
Mohamad El Abed	Allocataire MENESR			X	09/2008
Chrystel Gaber	CIFRE FT R&D	X			10/2010
Romain Giot	ANR ASAP			X	10/2010
Aude Plateaux	CIFRE BULL		X		09/2010
Johann Vincent	CIFRE SFR		X		01/2010

TABLE 5.2 – Liste des doctorants

Prénom et nom	Lieu	Thèmes			Date d'entrée
		T1	T2	T3	
Baptiste Hemery	ATER ENSICAEN		X		2010
Alexandre Ninassi	Post-doc CPER Sécurité		X		2010

TABLE 5.3 – Liste des ATER et post-doc

Prénom et nom	Type	Thèmes			Date d'entrée
		T1	T2	T3	
Mohamed Achemlal	PAST ENSICAEN		X		2007
Wilfried Aubry	Ingénieur R&D ENSICAEN	X			2007
Olivier Catherine	Ingénieur R&D (ADS+)	X			2010
Wipa Chaisantikulwat	Ingénieur R&D (RUNN)	X			2010
Julien Dary	Ingénieur d'étude ENSICAEN	X	X	X	2008
Sébastien Duval	Ingénieur R&D ENSICAEN	X			2008
Mohamad El Abed	Moniteur ENSICAEN			X	2008
Samuel Jolly	Ingénieur R&D ENSICAEN	X			2009
Philippe Lefebvre	PRAG ENSICAEN			X	2009
Julien Mahier	Ingénieur R&D ENSICAEN			X	2010
Jean-Claude Pailès	Ingénieur R&D (ANDROID, SFR)		X		2008
Latifa Rabhi	Ingénieur R&D (Caen la mer)	X			2010
Marie Reveilhac	Ingénieur R&D (PERSOPOLIS)	X			2008
Joan Reynaud	Ingénieur R&D ENSICAEN		X		2007
Jean-Jacques Schwartzmann	PAST ENSICAEN	X		X	2010
Sylvain Vernois	Ingénieur R&D ENSICAEN	X			2007

TABLE 5.4 – Liste des autres membres contractuels

5.1.2 Auto-évaluation

5.1.2.1 Points forts

Les points forts de cette équipe en émergence sont au nombre de quatre :

- Un positionnement thématique original dans le paysage national et international : Les travaux en monétique ont une visibilité internationale tant la France est reconnue dans ce domaine et Caen est un acteur incontournable national. Les travaux en informatique de confiance sont assez peu traités en France au grand dépit de l'Agence nationale de la sécurité des systèmes d'information (ANSSI) encourageant les activités dans ce domaine piloté par les USA au sein du Trusted Computing Group (TCG). La biométrie est malgré tout encore assez faiblement traitée en France. Un rapport de la DGRI (Direction Générale pour la Recherche et de l'Innovation) en 2007 sur la sécurité informatique a recensé seulement 7 enseignants-chercheur en France travaillant sur cette thématique. L'équipe aura fin 2010, 4 thèses sur ce thème et participe déjà aux meilleures conférences du domaine et au dépôt de projets européens sur ce sujet.
- Une valorisation industrielle excellente : Les relations industrielles dans le cadre de contrats, projets collaboratifs et thèses CIFRE sont un point fort incontestable de cette équipe. Le groupe de recherche dont est issue cette nouvelle équipe se positionne déjà comme l'un des plus actifs au sein du GREYC concernant les relations industrielles.
- La prise en compte des usages : La collaboration avec deux maîtres de conférences en psychologie perceptive permet la prise en compte des usages des utilisateurs de ces nouvelles technologies. Ceci est pour nous une composante importante de ces travaux.
- Une implication forte au sein du pôle de compétitivité TES : les membres de cette équipe participent très activement dans ce pôle (rédaction des enjeux scientifiques, comité de projets, groupement scientifique opérationnel, conseil d'administration...)

5.1.2.2 Points faibles

Les principaux points faibles sont au nombre de deux :

- Publication : Le faible nombre de permanents et notamment de docteurs sont des facteurs expliquant la faible culture de la publication. De nombreux ingénieurs R&D travaillant à l'ENSICAEN avant d'intégrer le GREYC ne publiaient pas dans le cadre de leurs projets collaboratifs de type FUI, ANR ou ITEA. L'amélioration de la qualité et de la quantité des publications est la priorité numéro 1 de l'équipe sur le prochain quadriennal.

- Nombre de permanents : L'activité de valorisation importante de l'équipe implique un nombre important de recrutement de contractuels. Il est nécessaire d'avoir un encadrement de titulaires qui sont trop peu nombreux dans l'équipe. Ceci est aussi un problème pour le co-encadrement de doctorants. Nous avons assez peu de difficultés pour trouver des candidats de qualité pour des thèses et les financements associés mais peu de doctorant parmi les 8 de l'équipe sont co-encadrés par un maître de conférences. L'arrivée d'un nouveau maître de conférences et d'un étudiant en post-doc en 2010 nous aidera incontestablement sur ce point. Afin de pallier ce problème, nous avons également demandé à deux maîtres de conférences de deux autres équipes du GREYC de co-encadrer deux de nos doctorants.

5.1.2.3 Opportunités

Les opportunités détaillées ici montrent comment l'équipe va lutter contre nos points faibles. Nous en listons quatre principales :

- Afin d'améliorer notre efficacité en terme de publication, nous avons recruté un grand nombre de doctorants, la plupart étant financés par des projets collaboratifs ou des thèses CIFRE. Afin d'améliorer notre visibilité nationale et internationale, les trois nouvelles thèses en septembre 2010 vont être co-dirigées par des laboratoires français et internationaux (Telecom ParisTech, Telecom SudParis, Université de Toronto au Canada). Ces co-directions nous permettront également de contourner le problème d'encadrement au sein de notre équipe et nous permettre de viser les publications de haut niveau.
- Nous avons l'habitude jusqu'à présent de recruter des ingénieurs R&D dans le cadre du CPER STIC et Sécurité. Nous avons gagné en maturité et nous envisageons de recruter désormais des post-doctorants sur ces supports. Un post-doc a été recruté pour la rentrée 2010.
- Un des nos objectifs est la participation à des projets internationaux sur le prochain quadriennal (ITEA, FP7, EUREKA, INTERREG, programmes internationaux ANR). Un projet Franco-Estonien a été déposé à l'EGIDE sur le calcul distribué dans un réseau d'objets communicants. Nous avons également participé à des propositions de projets de type FP7 en 2009 et 2010 mais ils n'ont pas été financés. Malgré tout, ces participations infructueuses nous permettent de nous insérer dans un réseau européen sur nos thématiques. Un des membres de l'équipe va participer au comité national de la normalisation en Biométrie piloté par l'AFNOR avec comme objectif la participation au comité international. Ceci renforcera notre visibilité nationale et internationale sur ces activités.
- Le faible nombre de permanents au sein de l'équipe peut être corrigé avec le temps tant les besoins d'enseignants sur nos thématiques sont importants à l'ENSICAEN. De plus, des ingénieurs R&D (auto-financés par la formation en monétique par apprentissage de l'ENSICAEN) pourront obtenir un contrat CDI dès lors que l'ENSICAEN sera passé aux compétences élargies (RCE) ce qui est prévu à échéance de 2012. Ceci ouvre des perspectives d'évolution positive du nombre de titulaires de l'équipe.

5.1.2.4 Risques

La valorisation industrielle est importante dans notre équipe. Un des risques dont nous sommes conscients est de rester à un niveau applicatif au détriment de la recherche fondamentale. Notre reconnaissance par les industriels du domaine nous permet généralement de négocier des thèses CIFRE avec une présence à 100% dans nos locaux (ce qui est le cas pour 3 des 5 thèses CIFRE). Ceci reste malgré tout une préoccupation importante de notre équipe.

5.1.3 Perspectives scientifiques

Les perspectives scientifiques communes aux trois thèmes de l'équipe concernent la sécurité et le respect de la vie privée. Ces propriétés d'une transaction électronique (paiement, transmission d'informations ...) souvent considérées comme contradictoires peuvent être traitées de façon conjointe.

5.1.3.1 Thème 1 : Monétique

La monétique est acutellement en pleine évolution notamment en Europe avec l'uniformisation des paiements prévue à l'échéance de 2012 (SEPA : Single Euro Payments Area). Ceci s'accompagne de l'utilisation de nouvelles technologies à base de XML, de réseaux IP (au lieu de réseaux privés) et de nouveaux terminaux générateur de la transaction (qui se limitait presque exclusivement auparavant à une carte à puce). L'essor des objets communicants (téléphone portable, iPad, GPS . . .) permet d'entrevoir de nouveaux services comme notamment le paiement sans contact (plus rapide), la diffusion d'informations culturelles ou touristiques, la fidélité, la billétique . . . Ceci pose des problèmes de sécurisation du service, de téléchargement de façon sécurisée sur l'objet communicant et d'accès à des informations personnelles. Ceci aura des impacts sur les applications, l'infrastructure de téléchargement, sur le back office de traitement des transactions et sur les protocoles d'échange de données (devant respecter la vie privée de l'utilisateur).

L'équipe forte de son expérience en monétique (notamment basée sur les cartes à puce) a pour objectif de contribuer à l'essor des ces nouveaux moyens de paiement. Les projets VACAMS et PERSOPOLIS nous donnent une très bonne compétence sur la post-personnalisation (téléchargement d'applications a posteriori) et la cohabitation sécurisée de plusieurs applications sur un objet communicant (carte multiservices étudiante, téléphone portable). L'utilisation des réseaux IP en paiement face à face permet aussi la convergence de solutions monétiques dans les différents contextes (paiement face à face, sur internet, compte bancaire en ligne). Le projet ADS+ a pour objectif de contribuer à cette généralisation au niveau du front office. Une thèse va débutée en septembre 2010 sur la définition de nouvelles architectures monétique (exploitant le téléphone portable) limitant la fraude bancaire avec Orange Labs.

5.1.3.2 Thème 2 : Informatique de confiance

Dans un monde de plus en plus numérique, de nombreuses transactions se font de façon électronique (paiement, déclaration aux impôts, e-commerce). Ceci pose des questions de sécurité (quand sera t-on capable d'assurer un niveau de sécurité comparable pour un paiement sur internet que lors d'une transaction en face à face?) et d'éthique vis à vis des de la protection des données personnelles d'un client. Le terminal générateur de la transaction peut prendre plusieurs formes (téléphone portable, PC, iPad...) et il est difficile de garantir la sécurité de celui-ci. Par ailleurs, la multitude d'identités numériques associées à un individu posent des problèmes de respect de la vie privée. A t-on vraiment besoin de donner sa date de naissance pour prouver que l'on a droit un tarif réduit pour un service quelconque?

Les solutions envisagées sont nombreuses pour securiser le terminal générateur de la transaction : techniques de 'trusted computing' (vérification au niveau du processeur de l'autorisation d'une action ou accès à une donnée), de vérification d'inocuité des programmes, de virtualisation des OS, moyens de rendre les interactions machine-utilisateur plus robustes. Afin de mieux respecter la vie privée des utilisateurs, il est possible d'utiliser des protocoles cryptographiques (du type *zero knowledge*) pour éviter le traçage d'un individu, etc. Une thèse va débuter en septembre 2010 sur cette thématique avec la société BULL afin de déployer ce type de protocoles et de montrer leurs intérêts opérationnels pour plusieurs applications (anonymisation de dossiers médicaux, paiement sur internet sans trace . . .).

5.1.3.3 Thème 3 : Biométrie

La biométrie est en plein essor et devrait être plus largement utilisée à l'avenir pour garantir l'authentification d'un individu pour une transaction électronique. La biométrie est souvent considérée par les académiques comme une application du traitement du signal et des images, de la reconnaissance de formes ou même de la cryptographie. Nous considérons pour notre part la biométrie comme un domaine à part entière au carrefour de plusieurs compétences. Les trois enjeux scientifiques majeurs de ce domaine auquel nous souhaitons répondre sont :

- Performance : La sécurité d'un système biométrique nécessite une bonne performance en terme d'erreur de vérification ou identification. Nous travaillons sur de nouvelles méthodes d'apprentissage statistique basées sur des architectures profondes (traitement conjoint de la sélection des attributs pertinents et de la séparation optimale) dans le cadre du projet ANR ASAP. Nous travaillons également sur la multimodalité consistant à combiner plusieurs modalités biométriques (avec des

approches originales basées sur la programmation génétique). La mise à jour du modèle biométrique est un élément essentiel notamment pour les modalités biométriques comportementales, nous utilisons également des algorithmes évolutionnaires pour traiter ce point. Nous considérons dans tous les cas des contraintes opérationnelles (notamment liée à la monétique) pour implémenter nos systèmes biométriques basées sur la dynamique de frappe au clavier, veines de la main, plis des doigts, visage ... Nous travaillons également sur le calcul de performance de systèmes biométriques sur des benchmarks en utilisant du cloud computing (collaboration avec l'Estonie).

- Sécurité : Avoir un algorithme performant n'est pas suffisant pour garantir la sécurité d'un système biométrique. Nous sommes en phase de développement d'une méthodologie permettant de qualifier la sécurité d'un système biométrique intégrant les choix technologiques et algorithmiques réalisés dans la conception d'un système biométrique. Notre objectif est de qualifier la sécurité d'un système biométrique par un indice mesurant sa vulnérabilité à différentes attaques. La qualification de la qualité de la donnée biométrique est un enjeu important pour limiter les attaques d'un système biométrique notamment pendant l'enrôlement. Nous avons défini dans le cadre d'une collaboration avec l'équipe Image du GREYC, une métrique générique de qualité biométrique combinant qualité d'image et qualité de l'information.
- Respect de la vie privée : Une information biométrique est une donnée sensible qu'il est impossible de révoquer. Nous travaillons sur des techniques algorithmiques (Biohashing en anglais) permettant la révocation d'une donnée biométrique. Un biocode (donnée biométrique révocable) est généré à partir d'un aléa et une donnée biométrique. Le biocode d'un individu est stocké et peut être régénéré à partir d'un autre aléa. Une étude est en cours en collaboration avec l'équipe Automatique du GREYC sur la robustesse du biocode généré (Peut on générer à l'infini des biocodes sans divulgation sur la donnée biométrique?). Il est également possible de stocker cette donnée révocable sur un passeport ou une carte à puce pour garantir sa protection. Ceci fait l'objet de travaux conjoints avec les deux premiers thèmes de l'équipe.

5.1.3.4 Aspects méthodologiques

L'équipe est structurée avec un responsable d'équipe ayant un rôle d'animation et de représentation au comité scientifique du laboratoire. Chacun des trois thèmes de recherche est animé par un membre de l'équipe. Une réunion d'équipe aura lieu en moyenne toutes les 6 semaines (rythme actuel du groupe Monétique & Biométrie). A l'occasion de ces réunions, une présentation de 20 minutes de chaque thème est réalisée concernant l'avancement d'une thèse, d'un projet ou d'une préparation d'une présentation à une conférence. Les aspects budgétaires, politique scientifique de l'équipe et les informations de la vie du laboratoire y sont également abordés. Les responsables de thème organisent à leur convenance des réunions thématiques sur l'avancement du thème.

Afin de compléter l'organisation de l'équipe, nous nous sommes dotés d'outils complémentaires. Nous avons mis en place un serveur subversion permettant d'avoir une démarche qualité sur le développement de nos logiciels et de la rédaction de nos articles en latex. Un serveur de suivi de bugs nous permet de renforcer cette démarche visant à exploiter au mieux le travail des stagiaires, ingénieurs et doctorants. Enfin, nous avons mis en place au sein de l'équipe un extranet permettant de fournir les informations utiles pour ses membres (publications, projets en cours, liens vers les conférences et revues pour les publications, transparents de présentation de l'équipe...).

Tous ces éléments garantissent une bonne diffusion de l'information et une émulation réciproque des membres de l'équipe. Cette dynamique engagée au sein du groupe Monétique & Biométrie a pour objectif de positionner cette nouvelle équipe au niveau international sur ses thématiques de recherche.

6

Équipe Image

6.1	Perspectives scientifiques	59
6.2	Projet 1 : Variationnel et multiéchelle sur les variétés	59
6.3	Projet 2 : Modélisation géométrique et topologique	60
6.3.1	Motivation	60
6.3.2	Verrous	60
6.3.3	Démarche	61
6.4	Projet 3 : Reconnaissance d’objets dans des images et des vidéos, indexation, classification.	61
6.5	Projet 4 : Ingénierie des connaissances pour la conception d’applications et l’évaluation en imagerie	62
6.5.1	Motivations	62
6.5.2	Verrous	62
6.5.3	Démarche	63
6.6	Vie de l’équipe	63
6.7	Risques	63
6.8	Opportunités	64
6.9	Politique d’intégration de non produisants	64
6.10	Références bibliographiques	64

Équipe Image

Responsable : Abderrahim ELMOATAZ

L'affectation des membres permanents de l'équipe Image aux nouveaux thèmes est détaillée ci-après.

Nom et Prénom	Fonction	Thèmes			
		P1	P2	P3	P4
Bougleux Sébastien	MCF		x		
Brun Luc	PR		x		
Chahir Youssef	MCF			x	
Charrier Christophe	MCF				x
Clouard Régis	MCF HDR				x
Condat Laurent	CR2	x			
Elmoataz Abderrahim	PR	x			
Fadili Jalal	MCF HDR	x			
Fourey Sébastien	MCF		x		
Janik Jean-Marie	MCF	x			
Jurie Frédéric	PR			x	
Lezoray Olivier	PR	x			
Mahboubi Amal	MCF		x		
Mokhtari Myriam	MCF		x		
Porquet Christine	MCF				x
Revenu Marinette	PR				x
Sophie Schupp	MCF	x			
Tschumperlé David	CR1	x			

6.1 Perspectives scientifiques

Au vu du bilan de l'équipe, tant quantitatif que qualitatif, les membres de l'équipe proposent une organisation pour le prochain quadriennal autour de quatre projets scientifiques. Les thématiques auxquelles ils se rapportent sont importantes pour la communauté. Les acteurs investiront leurs compétences reconnues et complémentaires afin d'œuvrer ensemble au service de ces projets scientifiques communs. Un travail de concertation a d'ores et déjà été mené entre les membres de l'équipe, afin d'identifier les questions ouvertes tant sur le plan théorique qu'appliqué. Les participants de chaque projet ont cerné les problèmes ouverts et les verrous à lever, qui constitueront autant de défis pour les prochaines années.

6.2 Projet 1 : Variationnel et multiéchelle sur les variétés

Motivation. Les imageurs modernes fournissent des données de plus en plus complexes, dont les échantillons sont acquis sur des grilles non nécessairement cartésiennes équi-espacées et ne sont pas forcément à valeurs scalaires. En plus des données vectorielles que sont les images couleurs, multi-spectrales, ou multi-modalités, nous assistons à l'apparition de données très riches dont les valeurs structurées obéissent à des contraintes non linéaires. On peut voir ces données comme une application $f : \Omega \rightarrow \mathcal{V}$, où \mathcal{V} est une variété. Des exemples typiques de \mathcal{V} sont S^1 ou S^2 (données de direction, polarisation), $SO(3)$ (données d'orientation 3D), les matrices définies-positives (typiquement en IRM de diffusion), les variétés grassmaniennes (données projetées dans un sous-espace de \mathbb{R}^d). En outre, de nombreux problèmes en traitement d'image reposent sur un traitement non local de l'information, ce qui peut se traiter en considérant \mathcal{V} comme un espace de patches. Pour le domaine Ω , outre les treillis de \mathbb{R}^d , les exemples incluent la sphère, les sections coniques, les surfaces, ou encore le cas important des graphes.

Verrous à lever. Des efforts particuliers devront être fournis vis-à-vis des points durs suivants :

- Des constructions de la transformée en ondelettes ont été proposées pour certaines variétés non Euclidiennes Ω dont la sphère, les variétés compactes différentiables, les variétés lisses (ondelettes de diffusion), ou encore les graphes. Malgré des travaux prometteurs [5], beaucoup reste à faire dans ce domaine.
- L’analyse par ondelettes dans le cas où Ω est une grille Cartésienne, mais où \mathcal{V} est une variété quelconque a été considérée dans la littérature par très peu d’auteurs, dont [7]. Leur représentation multiéchelle s’appuie sur le schéma de raffinement (par interpolation) de Deslauriers-Dubuc, agissant dans le plan tangent de la variété.
- La formulation des problèmes de traitement de données implique dans beaucoup de situations de minimiser des fonctions objectives non lisses ou non convexes. Malgré l’intérêt considérable suscité notamment par l’analyse convexe, l’optimisation non lisse [1] ou non convexe demeure peu développée. Des algorithmes rapides, avec les garanties théoriques de convergence, restent à concevoir.

Démarche. Afin d’aborder les thématiques de ce projet, les participants s’appuieront sur leurs compétences en approches variationnelles, EDP et EdP sur les variétés, représentations parcimonieuses multiéchelles, analyse harmonique, optimisation, théorie de l’échantillonnage, et problèmes inverses. Ils continueront par ailleurs leur investissement dans différents domaines d’application, en particulier autour de l’imagerie (bio-)médicale et astronomique. Quelques perspectives qui nous apparaissent pertinentes à explorer sont les suivantes :

- La conception d’analyses multiéchelles (autres que les ondelettes) ou de schémas de raffinement pour des fonctions définies sur des variétés à valeurs dans des variétés, comme les ensembles de patches ou les données de polarisation sur la sphère en imagerie astronomique. Cet axe nécessitera des compétences théoriques, parmi lesquelles la théorie des groupes et algèbres de Lie, la théorie des matrices aléatoires, ou encore l’analyse spectrale sur graphes. Aussi, des compétences de mise en œuvre et d’implémentation rapide sont requises.
- À l’aide de ces méthodes d’analyse adaptées aux variétés, nous nous attaquerons à des problèmes de traitement de données correspondants, comme la reconnaissance, la détection, la segmentation, et les problèmes inverses tels que la restauration et la reconstruction. En particulier, la photographie computationnelle connaît un essor considérable avec l’avènement dans nos sociétés du « tout numérique ». La capture d’images couleurs 2D ou 3D et tous les traitements post-acquisition reposent sur des techniques modernes de traitement du signal et des images. Dans ce vaste domaine, les méthodes d’analyse et d’optimisation sont d’un grand secours pour la conception conjointe et robuste des processus de reconstruction et restauration requis.

6.3 Projet 2 : Modélisation géométrique et topologique

6.3.1 Motivation

Le caractère discret de la réalité pressenti par Démocrite impacte notre réalité quotidienne où nous devons manipuler des ensembles finis ou dénombrables d’objets en relations. Bien que beaucoup de phénomènes puissent être approximés par les mathématiques continues, de nombreux objets restent intrinsèquement discrets. Les formalismes mathématiques permettant de décrire ces objets varient en fonction des espaces dans lesquels ils sont plongés : La géométrie discrète a pour objet l’étude des propriétés géométriques et topologiques d’objets définis sur des pavages réguliers. La géométrie algorithmique étudie quand à elle les propriétés d’un ensemble discret de primitives géométriques plongées dans un espace continu. Les objets définis comme des partitions d’espaces de type pseudo-variété non-orientable (ou orientable) sont quant à eux efficacement décrits par le formalisme des cartes combinatoires. Enfin, la théorie des graphes décrit les objets ne présentant pas de plongements spécifiques.

L’objet de ce thème est la modélisation, le traitement et la mesure de similarité entre objets discrets à l’aide des outils mathématiques décrits précédemment.

6.3.2 Verrous

La reconnaissance d’objets discrets se heurte depuis la fondation de cette discipline à deux verrous importants et intimement liés :

6.4. PROJET 3 : RECONNAISSANCE D'OBJETS DANS DES IMAGES ET DES VIDÉOS, INDEXATION, CLASSIFICATION

1. Les espaces d'objets discrets ont généralement peu de propriétés. Une des principales lacunes de ces espaces est la notion de distance entre objets discrets qui ne définit généralement pas une métrique. Cette lacune ferme aux objets discrets des pans entiers des mathématiques dont de nombreuses méthodes de classification. Calculer la moyenne d'un ensemble de graphes est un exemple de problème difficile du fait du faible nombre de propriétés de ces espaces.
2. La modélisation discrète d'objets bruités implique un niveau d'abstraction supplémentaire via la notion de classe afin de capturer la notion de variabilité intra classe. Dans le cas continu, une classe est souvent modélisée par une distribution. Ce type de phénomène est pour l'instant peu modélisé [8] dans le cas discret ce qui rend délicat la définition de classe d'objets à partir, par exemple de modèles statistiques. Notons toutefois que la définition d'une métrique sur un espace discret est un grand pas vers la définition des modèles statistiques.

6.3.3 Démarche

Nous comptons étudier les notions d'espaces discrets définis à partir de différentes métriques. Une des approches que nous comptons privilégier est la notion de noyaux sur graphes. Les méthodes de noyaux sur graphes apparues récemment permettent un plongement des objets discrets dans des espaces de Hilbert, ouvrant ainsi l'ensemble des méthodes statistiques et variationnelles aux espaces discrets. Notons toutefois que cette ouverture se paye encore par une mesure de similarité qui reste généralement moins précise que les mesures issues de méthodes d'appariement. Dans le cadre des graphes, nous comptons approfondir les propriétés des noyaux en étudiant notamment la définition de noyaux à base de motifs complexes et la notion de distance entre sacs de motifs. Des applications en chémoinformatique et en reconnaissance de formes sont envisagées.

Dans le cadre de la reconnaissance d'objets 3D, nous comptons traiter le problème du bruit et de la distance intra-classe à l'aide d'une segmentation ascendante hiérarchique de surfaces polygonales 3D basées sur des critères énergétiques et des coupes optimales dans des hiérarchies [4] (à l'image de ce que nous avons déjà réalisé dans le cadre de la segmentation d'images [6]). Une première étape consistera à mettre en place un modèle en utilisant des critères énergétiques sans a priori sur les primitives géométriques composant les objets. La seconde étape consistera à étendre ces travaux pour décomposer une surface polygonale en un ensemble hiérarchique de primitives. Enfin, nous pensons lier ces travaux au premier projet en étudiant des méthodes d'appariement de formes basées sur la décomposition hiérarchique de celles-ci induite par la segmentation. Nous étudierons deux approches duales : des noyaux sur graphes définis directement sur la segmentation hiérarchique de la surface polygonale et des noyaux définis sur les simplifications successives du squelette initial induit par la segmentation. Ce projet combine les compétences des acteurs en géométrie algorithmique, modélisation topologique et segmentation.

6.4 Projet 3 : Reconnaissance d'objets dans des images et des vidéos, indexation, classification.

Motivations. Bien qu'ayant fait l'objet de nombreuses recherches, la reconnaissance d'objets dans des images (qu'elles soient fixes ou animées) reste à ce jour un défi important. Ce champ de recherche englobe différentes tâches de vision comme (a) la *détection* d'objets dans des images (prédire la position des objets de catégories connues (véhicules, piétons, animaux, etc.)), (b) la *segmentation de catégories d'objets* (extraire les pixels des images qui appartiennent à des classes d'objets connues et modélisées), (c) la *classification de catégories d'images* (affecter à des images des labels en accord avec les objets contenus dans les images) (d) la *recherche et l'indexation d'images ou de vidéos* (enrichir les images d'annotations qui décrivent leur contenu et ainsi de permettre leur recherche dans des bases). Les travaux que nous développerons dans ce projet visent à apporter des contributions à ces différentes tâches de vision.

Verrous. (a) Apprentissage. La construction de modèles d'objets ou de scènes nécessite d'avoir des volumes importants de données d'entraînement. Certains mécanismes ingénieux (comme le site <http://labelme.csail.mit.edu/>) permettent de récupérer des annotations construites de manière collective mais cela reste insuffisant. L'exploitation d'annotations fournies par des experts, leur propagation, leur diffusion sont également des sujets difficiles sur lesquels des progrès doivent être faits. (b) Représentation

des images. La représentation des images est une étape clé des algorithmes d'interprétation d'images. Beaucoup de progrès ont été faits durant les dernières années, que ce soit au niveau des représentations locales (descripteurs SIFT, FERNS, LBP, etc.) ou que ce soit au niveau de la représentation de régions d'images ou d'images (modèles de type sac-de-mots, modèles PLSA, LDA, etc.). Malgré ces progrès, ces modèles en question montrent assez vite leurs limites ; la prise en compte des relations spatiales demeure par exemple insatisfaisante ; de même les variations de pose ou d'apparence, lorsque l'on traite des catégories d'objets, sont insuffisamment prises en compte. (c) Saillance. Plusieurs modélisations de saillances spatiales, visuelles et temporelles ont été proposées dans la littérature à partir des couleurs et à partir du mouvement. Des travaux récents proposent l'utilisation des cartes de saillance spatio-temporelle en reconnaissance d'images et en analyse vidéo. Cependant, peu de travaux ont été abordés dans le cadre des approches non locales ou de la théorie spectrale des graphes. (d) Apprentissage de distance. Dans de nombreux contextes, il est important de posséder une bonne mesure de similarité entre descripteurs (qu'il s'agisse de descripteurs locaux ou de descripteurs d'images). L'apprentissage de distances à partir de paires d'images (ou régions) similaires est une stratégie intéressante, mais qui se heurte à différents problèmes comme en particulier : le choix des attributs visuels à utiliser, la définition des espaces de représentation, l'optimisation des contraintes.

Directions de recherche. (a) *Apprentissage semi-supervisé* : si les données annotées sont rares et coûteuses à obtenir, les données non annotées sont en revanche abondantes (sites de partage de photos ou de vidéos par exemple). La combinaison d'une petite quantité de données annotées avec une grande quantité de données non-annotées nous semble une direction intéressante. Nous comptons adresser cette question de l'apprentissage semi-supervisé par des méthodes reposant sur des instances multiples et sur l'utilisation de sacs de motifs émergents dans un espace réduit. La question de la détection d'outliers est également intéressante et une piste qui nous semble intéressante est celle de l'utilisation de noyaux en prenant en compte le voisinage des instances. (b) *Représentation des images* : nous comptons poursuivre nos recherches sur l'apprentissage de codages locaux des images (vocabulaires visuels) adaptés à des catégories d'objets (en particulier adaptés à la description des visages et des personnes). Une des directions que nous souhaitons suivre est la représentation des images (ou des régions d'images) au moyen de sous espaces sémantiques, c'est-à-dire d'espace de projection dont les dimensions correspondent à des traits sémantiques des images. (c) *Saillance* : nous souhaitons traiter cette problématique dans un cadre unifié qui utilise des caractéristiques intra et interséquences qui exploitent la dynamique et la cinématique des enveloppes visuelles 2d+t (volumes composés d'actions unitaires) avec des mesures de similarités appropriées telles que la déformation temporelle dynamique. (d) *Apprentissage de distance* : nous souhaitons poursuivre les directions de recherches, à savoir l'utilisation de méthodes à base de forêts aléatoires, qui nous ont permis de faire des avancées intéressantes dans ce domaine.

6.5 Projet 4 : Ingénierie des connaissances pour la conception d'applications et l'évaluation en imagerie

6.5.1 Motivations

La complexité du développement de logiciels spécialisés en traitement d'images ou en vision freine leur réelle pénétration dans l'industrie. Les recherches en vision cognitive [2] visent à concevoir des systèmes de traitement d'images ou de vision plus adaptatifs, tenant compte des besoins spécifiques des utilisateurs. Dans ce cadre, nous nous intéressons à la conception de systèmes capables de générer automatiquement des logiciels de traitement d'images ou de vidéos configurés pour un objectif donné, non connu à l'avance. L'enjeu est de permettre à des utilisateurs non spécialistes du traitement d'images de produire eux-mêmes des logiciels adaptés à leurs besoins.

6.5.2 Verrous

Si cet objectif n'est pas nouveau, la conception de tels systèmes est toujours considérée comme un problème particulièrement complexe au sens de la systémique [10]. Le développement de logiciels de traitement d'images repose en effet sur une démarche empirique, faite de connaissances tacites, c'est-à-dire enracinées dans l'action et non-verbalisables. Cette particularité rend difficile le processus d'acquisition

des connaissances nécessaires à la conception du système. Les difficultés naissent aussi de l'absence de mesure universelle de la qualité d'images. La mesure de qualité d'une image ou d'une vidéo est une information précieuse qui peut être exploitée, soit de façon prospective pour sélectionner les meilleurs algorithmes de traitement parmi une base prédéfinie pour construire une chaîne de traitement, soit de façon rétrospective pour adapter dynamiquement le comportement d'une chaîne de traitement donnée.

6.5.3 Démarche

Forts de notre expérience en matière de représentation et d'exploitation de connaissances en traitement et évaluation d'images, les perspectives de recherche du projet s'articulent autour de deux axes thématiques complémentaires :

- **La définition de mesures de qualité d'images et de vidéos.** Deux types d'étude sont envisagées ici. La première étude porte sur la modélisation des critères de jugements humains pour définir une mesure de qualité de vidéos compressées sans référence. L'originalité de l'approche tient à la prise en compte d'informations non seulement sur les distorsions spatiales mais aussi sur les distorsions temporelles (rarement prises en compte jusqu'à présent). La seconde étude est dédiée à la modélisation des critères de jugements objectifs permettant de mesurer la qualité d'un résultat par rapport à un but d'utilisation. Ce travail original est rendu possible parce que nous disposons, par nos travaux précédents, d'un modèle de représentation des objectifs de traitement formalisé par une ontologie.
- **La construction automatique de chaînes de traitements d'images adaptées à un objectif donné par réutilisation d'expériences.** Le raisonnement à partir de cas se propose de résoudre des problèmes en s'inspirant de situations semblables déjà rencontrées et pour lesquelles on dispose de solutions [3]. Ce modèle permet d'éviter le goulot d'étranglement de l'acquisition des connaissances, en permettant l'intégration dans la base de cas de solutions « brutes » données par des spécialistes.

Notre originalité tient à l'exploitation de notre modèle de représentation d'objectif pour stocker chaque nouvelle solution fournie par un spécialiste avec la formulation de l'objectif correspondante. La recherche d'une solution s'envisage par comparaison entre l'objectif courant et les objectifs associés aux cas stockés dans la base, et devra procéder par des cycles d'interaction entre l'utilisateur et le système [9].

6.6 Vie de l'équipe

Chaque thème organisera ses activités autour de groupes de travail réguliers. La vie de l'équipe sera alimentée par des contacts quotidiens entre ces membres et par les séminaires de l'équipe Image qui sont organisés un à deux fois par mois. Notons enfin que si ces thèmes structurent le travail de l'équipe, ils ne constituent en rien des barrières étanches et que des objets tels que les graphes ou des champs disciplinaires tels que l'optimisation sur des variétés sont communs à plusieurs thèmes.

6.7 Risques

La faible attractivité du master LID¹ parcours image ne permet pas à l'équipe de recruter efficacement de nouveaux doctorants. Durant l'année 2009-2010, un seul étudiant de master 1 a choisi le master LID. Ce dernier point est à rapprocher au faible nombre (2) de membres de l'équipe Image présents à l'UFR science. Cette faible présence à l'UFR permet difficilement de sensibiliser les étudiants aux problématiques abordées dans le master LID. La majeure Image de l'ENSICAEN voit également ses effectifs décroître et est composée pour 2009-2010 de 17 étudiants inscrits en 3^e année et 14 en 2^e année. Nous pensons rapidement créer un club de partenaires industriels permettant une meilleure visibilité des métiers connectés à l'imagerie et ses applications.

Des recherches en Biométrie étant réalisées à la fois au sein du thème « Extraction et gestion des connaissances » de l'équipe Image et dans le groupe Monétique, on peut craindre une diminution de la visibilité de l'équipe Image et une perte de cohérence du discours du GREYC en région concernant

1. Langage, Image, Document

les compétences qu'il peut mobiliser dans le cadre du pôle de compétitivité Transaction Électroniques Sécurisées (TES). En effet deux équipes d'un même laboratoire se réclament actuellement des mêmes compétences. Le regroupement de cette activité dans une même équipe permettrait de résoudre cette ambiguïté.

6.8 Opportunités

La nouvelle structuration de la recherche en région représente des opportunités intéressantes pour l'équipe Image. Celle-ci compte tout d'abord poursuivre son implication dans le domaine de la santé notamment via sa participation à l'IRCBN². Cet institut a pour vocation de structurer l'ensemble de la recherche sur le Cancer en Basse-Normandie. L'équipe est fortement impliquée dans cette thématique notamment au travers de ses activités en micro-imagerie. Notons sur ce point que l'équipe Image dispose du seul microscope à lames entières de Basse-Normandie sur le site universitaire de Saint-Lô.

L'équipe compte également poursuivre ses activités concernant la sécurité par l'image et s'appuiera pour cela sur le pôle de compétitivité TES³.

L'équipe désire enfin s'investir dans les activités liées au document numérique en Basse-Normandie et renforcer ses compétences dans le domaine du traitement de maillage. Elle bénéficiera pour cela du soutien de l'association NovImage dont l'ENSICAEN et l'Université sont membres au travers de Luc Brun et Abder Elmoataz.

6.9 Politique d'intégration de non producteurs

L'effectif de l'équipe pour le prochain quadriennal comprend cinq enseignants chercheurs non producteurs en termes de recherche et valorisation. Ces 5 collègues ont manifesté leur volonté d'appartenir à l'équipe, ils se sont également engagés à contribuer et participer à la vie scientifique de celle-ci.

Cette intégration sera associée à un accompagnement leur permettant de devenir producteurs au cours du quadriennal prochain. Un des points forts de cet accompagnement se traduit par l'existence de 3 professeurs référents : L. Brun, M. Revenu, A. Elmoataz. Chacun de ces professeurs va suivre individuellement plusieurs non producteurs (L. Brun (2), A. Elmoataz(2), M. Revenu(1)) afin de faciliter leur intégration.

6.10 Références bibliographiques

- [1] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Technical Report R.I. 685, Centre de Mathématiques Appliquées (CMAP), Palaiseau, May 2010.
- [2] P. Auer et al. A research roadmap of cognitive vision. Technical Report v5, ECVISION, www.ecvision.org, 2005.
- [3] M. Frucci, P. Perner, and G. Sanniti di Baja. Case-Based-Reasoning for Image Segmentation. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 22(5) :829–842, 2008.
- [4] Laurent Guigues, Jean Pierre Cocquerez, and Hervé Men. Scale-sets image analysis. *Int. J. Comput. Vision*, 68(3) :289–317, 2006.
- [5] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 2009. soumis.
- [6] Jean Hugues Pruvot. *Segmentation et appariement hiérarchiques basés sur les pyramides combinatoires*. PhD thesis, Université de Caen Basse Normandie - ED SIMEM, France, 2008.
- [7] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4) :1201–1232, 2005.

2. Institut Régional contre le Cancer de Basse-Normandie

3. Transactions Electroniques Sécurisées

- [8] Sébastien Rebecchi. *Formalisme statistique pour ensembles de structures discrètes*. Thèse de doctorat en informatique, Institut National des Sciences Appliquées de Lyon, September 2009.
- [9] S. Santini, A. Gupt, and R. Jain. Emergent Semantics through Interaction in Image Databases. *IEEE Trans. Knowledge and Data Engineering*, 13(3) :337–351, 2001.
- [10] D. Vernon. Cognitive vision : The case for embodied perception. *Image and Vision Computing*, 26(1) :127–140, 2007.

7

Équipe Électronique

7.1	Equipe Electronique	67
7.1.1	Auto-évaluation	67
7.1.1.1	Opportunités	67
7.1.1.2	Risques	67
7.1.2	Perspectives scientifiques	68
7.1.2.1	Thème 1 : « Microcapteurs à base d'oxydes fonctionnels & Bruit 1/f »	68
7.1.2.2	Thème 2 : Magnétométrie à très haute sensibilité - du capteur à l'instrument.	69

Équipe Électronique

Responsable : Christophe DOLABDJIAN

7.1 Equipe Electronique

Prénom et nom	Fonction	Thème principal	Thème secondaire
Régis CARIN	Pr.	T1	
Christophe DOLABDJIAN	Pr.	T2	
Stéphane FLAMENT	Pr.	T1	T2
Didier ROBBES	Pr.		
Jean Marc ROUTOURE	Pr.	T1	
Jean BAYARD	Pr.	Site d'Alençon	
Laurence MECHIN	CR(HDR)	T1	
Marc LAM CHOK SING	MC	T2	T1
Chantal GUNTHER	MC	T1	T2
Christophe CORDIER	MC	T2	
Sébastien SAEZ	MC	T2	
Bogdan CRETU	MC	T1	
Bruno GUILLET	MC	T1	
Matthieu DENOUAL	MC	T2	
Gilles ALLEGRE	MC	T2	
Corentin JOREL	MC	T2	T1
Pierre LANGLOIS	MC	T1	

7.1.1 Auto-évaluation

Pour ce nouveau quadriennal, l'équipe a projeté de s'organiser en deux thèmes. Les thèmes 1 et 2 s'intitulent, respectivement, : « Microcapteurs à base d'oxydes fonctionnels & Bruit 1/f » et « Magnétométrie à très haute sensibilité - Du capteur à l'instrument. ». Cette nouvelle organisation permet à l'équipe d'afficher distinctement ses deux activités principales qui portent respectivement sur **l'étude de composants** (principalement à base de LSMO ou semi-conducteurs) et de leur bruit à basse fréquence, et **l'étude de capteurs à très haute sensibilité** (principalement magnétique) afin d'appréhender au mieux la problématique de mise en oeuvre d'instruments. Elle poursuivra ainsi, au sein de ces thèmes, ses recherches dans la conception, la réalisation, l'étude et la mise en oeuvre de dispositifs.

7.1.1.1 Opportunités

- S'inscrire dans les groupements des laboratoires européens travaillant sur les oxydes fonctionnels
- Organisation ou partenaire dans l'organisation de la conférence ICNF en 2013 (avec l'IES de Montpellier)
- Organisation de la 4ème édition de l'école thématique internationale du CNRS "High sensitivity magnetometers : Sensors & Applications"
- Renforcer nos liens et développer des projets avec les acteurs des pôles de compétitivité (TES & MOVE'O), notamment autour de la mécatronique et des systèmes embarqués

7.1.1.2 Risques

Les activités parfois prospectives (fabrication de micro-capteurs, développement d'applications...) reposent sur des aléas technologiques et/ou des difficultés de mise en oeuvre qui peuvent mener à des écueils.

7.1.2 Perspectives scientifiques

7.1.2.1 Thème 1 : « Microcapteurs à base d'oxydes fonctionnels & Bruit 1/f »

Responsable : L. MECHIN

Mots clés : Micro-capteurs, MEMS & NEMS (Micro- & NanoElectroMechanical Systems), Bruit en 1/f, Couches minces, Oxydes fonctionnels, Spintronique, Microélectronique, Silicium, Bolomètre, Magnétorésistance, Imagerie magnétique, Domaines magnétiques

Introduction/contexte : Les activités de recherche sur les oxydes fonctionnels vont continuer à porter principalement sur le développement de micro-capteurs en couches minces de manganites $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (LSMO). Le LSMO présente plusieurs propriétés intéressantes pour la réalisation de micro-capteurs : la variation de résistance lors de la transition métal – isolant, les dépendances de ses propriétés structurales, électriques et magnétiques avec les contraintes et la polarisation en spin proche de 100%. Nous chercherons notamment à utiliser ces propriétés pour la réalisation de détecteurs de rayonnement ou de thermomètres (axe 1), de composants de type MEMS & NEMS sur silicium afin de renforcer les effets de contraintes en introduisant des matériaux piézoélectriques (axe 2) et des composants spintroniques planaires (axe 3).

Les perspectives de recherche en ce qui concerne le bruit et les fluctuations à basse fréquence (axe 4) s'inscrivent dans la poursuite des activités actuelles :

- comprendre l'origine du bruit de façon à concevoir et réaliser des composants et capteurs plus performants car moins bruyants. L'étude du bruit permet également de fournir des modèles compacts pour les outils de simulation,
- évaluer la qualité de réalisation des dispositifs car le bruit à basse fréquence est un facteur de mérite très sensible,
- dans une approche plus fondamentale, comprendre les mécanismes associés aux transports des charges électriques car à chaque mécanisme de transport correspond une source de bruit basse-fréquence qu'il convient d'identifier par des études en température par exemple.

AXE 1 : Détecteurs de rayonnement et thermomètres à base de manganites Nous avons démontré que les couches minces de $\text{La}_{0.7}\text{Sr}_{0.3}\text{MnO}_3$ (LSMO) présentent des propriétés encourageantes pour les applications en thermométrie et en bolométrie non refroidie (Noise Equivalent Power < 1 pW.Hz-1/2 à 300 K). Dans le prochain quadriennal, notre objectif est d'étudier de façon systématique les effets des interactions lumière-manganites en fonction notamment de la longueur d'onde et de l'intensité de la source laser utilisée. Des effets thermiques, mais aussi photoconductifs, seront considérés. Nous étudierons notamment les effets du type de substrat sur la sensibilité de la réponse et sur la constante de temps de réponse. De plus, nous savons déposer les couches minces de LSMO sur silicium, ce qui permet d'envisager deux familles de détecteurs :

- des structures P-I-N utilisant la conduction de type P dans LSMO et de type N dans le silicium,
- des microbolomètres LSMO suspendus. Les premières fabrications de structures suspendues ayant été réalisées, des modélisations électro-thermiques couplées aux mesures permettront de déterminer les meilleures dimensions des surfaces de détection afin d'optimiser la réponse bolométrique de ces détecteurs.

Ces travaux font partiellement l'objet de 2 thèses démarrées en oct. 2009 (S. Liu & A. Aryan), et d'une collaboration avec l'univ. de Cornell (USA).

En parallèle, nous étudierons les potentialités d'autres matériaux, par exemple une autre composition de $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ou un mélange $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3 / \text{La}_{1-x}\text{Ca}_x\text{MnO}_3$ présentant un coefficient de résistance en température le plus élevé possible à 300 K tout en conservant un faible niveau de bruit. Dans ce but, nous poursuivrons notre collaboration avec l'univ. de Naples (Italie) et nous avons initié début 2010 une collaboration avec l'univ. de Stuttgart (Allemagne).

AXE 2 : MEMS & NEMS à base d'oxydes fonctionnels L'utilisation des techniques de micro-usinage du substrat de silicium permet d'envisager une large variété de composants de type MEMS combinant manganites et matériaux piézoélectriques, tirant ainsi parti de la dépendance des propriétés des manganites aux contraintes. Nous proposons alors de réaliser des structures de type bicouches piézoélectriques/manganites sur des membranes de silicium micro-usinées afin d'accentuer ces effets et

de réaliser ainsi des capteurs de contraintes ou magnéto-électriques (pour lesquels, soit les propriétés magnétiques varient en fonction de tension appliquée, soit la résistivité varie en fonction d'un champ magnétique appliqué). A court terme, nous allons porter nos efforts sur l'optimisation des propriétés des couches piézoélectriques épitaxiées de BaTiO₃, PZT et sur la réalisation d'électrodes épitaxiées à base de SrRuO₃. Lorsque les couches piézoélectriques seront obtenues, nous devrons commencer à définir différentes géométries permettant d'appliquer des déformations connues aux couches de manganites LSMO et de mesurer les effets sur leurs propriétés. Ces travaux font partiellement l'objet d'une thèse démarrée en oct. 09 (S. Liu), d'une collaboration avec l'univ. de Cornell (USA).

De plus, une collaboration naissante a été initiée avec le LAAS à Toulouse avec pour objectif l'amélioration des propriétés des MEMS et NEMS piézoélectriques grâce à l'utilisation de matériaux piézoélectriques épitaxiés réalisés par ablation laser au GREYC. Si cet objectif est atteint, une grande gamme de composants pourra être imaginé et testé.

AXE 3 : Composants spintroniques planaires : Résistance et manipulation par transfert de spin de parois de domaine magnétique dans LSMO

Nous avons montré que les substrats vicinaux permettent d'induire une anisotropie magnétique uniaxiale dans une couche mince d'un matériau ferromagnétique tel que LSMO, et que les marches du substrat induisent des effets magnétorésistifs augmentés à champ magnétique faible. Nous proposons alors d'associer des effets de géométrie à l'utilisation de substrats vicinaux afin de créer des parois de domaines magnétiques dans LSMO à des endroits souhaités, donc en nombre contrôlé. L'objectif sera ensuite d'étudier la résistance de plusieurs parois de domaine magnétique, puis d'une seule paroi, ainsi que leur manipulation par transfert de spin. Notre but final est la réalisation de composants spintroniques planaires à base de LSMO travaillant à la température ambiante. Des structures optimisées de capteurs de champ magnétique seront proposées. Dans ce programme, outre la microscopie magnétique, l'apport des mesures de bruit, qui constitue une très forte originalité des travaux du GREYC, sera un outil essentiel afin de démontrer la présence, le piégeage ou le dépiégeage de parois de domaines magnétiques. En 2010, nous faisons l'acquisition d'un bâti de mesures sous pointes refroidi (10 – 400K) et sous champ magnétique (0,5 T) qui permettra de faire ces caractérisations fines.

AXE 4 : Bruit et fluctuations dans les composants actifs et passifs. On propose pour le prochain quadriennal de poursuivre les travaux en cours :

- sur les transistors MOS nanométriques en collaboration avec IMEC (Louvain Belgique),
- sur les couches minces d'oxyde de manganèse et autre composition d'oxyde fonctionnel dans le cadre des échanges internationaux du thème,
- sur les semiconducteurs III/V notamment les nitrures d'indium dans le cadre de l'ITN RAINBOW et de l'ANR COSNI en collaboration avec le CIMAP (Centre de Recherche sur les Ions, les Matériaux et la Photonique, UMR CNRS,CEA,UCBN,ENSICAEN 6252)

De façon plus prospective, on prévoit également d'étendre la gamme de composants aux composants de puissance de façon à renforcer la collaboration débutée dernièrement avec le LUSAC (Laboratoire Universitaire des Sciences Appliquées de Cherbourg). Un travail sur l'instrumentation de polarisation de ces dispositifs sera certainement à mener pour permettre une polarisation faible bruit ne nécessitant pas de batteries.

7.1.2.2 Thème 2 : Magnétométrie à très haute sensibilité - du capteur à l'instrument.

Responsable : S. SAEZ

Mots clés : Capteurs Magnétique, Magnétomètre, Magnéto-Impédance à effet Géant, Magnéto-Electrique, Magnétorésistance, Bolomètre, Substitution Electrique, Capteur intelligent, Faible bruit, Instrumentation, Imagerie magnétique, Contrôle non destructif, Biomagnétisme.

Introduction/contexte : Il y a plus d'une dizaine d'années, par une approche systémique des mesures magnétiques, nous avons fait le constat qu'il était parfois abusif de prôner l'utilisation de SQUID dans certaines applications en magnétométrie à très haute sensibilité. Depuis, nous oeuvrons à optimiser des systèmes de mesures magnétiques pour des applications spécifiques. Pour cela, nous cherchons d'une part à améliorer les performances des capteurs magnétiques et d'autre part, à avoir une très bonne

compréhension de la problématique de la mesure. Cette démarche de recherche a très largement porté ses fruits, tant en terme de résultats scientifiques que de collaborations et reconnaissances internationales. Les perspectives actuelles, en magnétométrie, s'inscrivent donc pleinement dans la continuité de ces travaux. Nous pouvons noter que des problématiques similaires sont apparues dans le cas de la mise en oeuvre de bolomètres. Ainsi, une démarche systémique, semblable à celle menée en magnétométrie, doit permettre de développer des synergies et conduire à des avancées significatives dans le domaine de la bolométrie.

Ainsi, il est proposé, pour les quatre ans à venir, de poursuivre nos travaux de recherche en magnétométrie. A court terme, notre expérience dans l'étude systémique des magnétomètres à haute sensibilité permettra de structurer une partie des activités en bolométrie. A plus long terme, il pourra être envisagé de développer des applications complémentaires, notamment dans le domaine du contrôle non destructif.

AXE 1 : Étude de capteurs magnétiques L'étude et le développement de capteurs magnétiques à très haute sensibilité, ayant un bruit intrinsèque d'environ $100 \text{ fT}/\sqrt{\text{Hz}}$ à 1 Hz, une large bande passante pour un volume actif effectif inférieur à un cm^3 , de résolution spatiale avoisinant le millimètre et fonctionnant à température ambiante, reste l'un des enjeux majeurs de nos recherches.

Nos activités antérieures nous ont permis d'identifier deux dispositifs prometteurs : les dispositifs à effet magnéto-inductif géant (GMI) et ceux à effet magnéto-électrique (ME). Il s'agit pour nous de poursuivre leur étude et d'en tirer le meilleur parti pour le développement de magnétométrie.

A court terme, la thèse de Basile Dufay devrait aboutir à la mise en oeuvre d'un magnétomètre à GMI. La collaboration avec l'École Polytechnique de Montréal devrait se poursuivre en partenariat avec une entreprise canadienne afin d'utiliser ces capteurs pour des applications dans le domaine du bio-médical. Notre collaboration avec Virginia Tech. dans le cadre du contrat " Heterostructural Uncooled Magnetic Sensor (HUMS) " avec la DARPA (sept. 2009 – sept. 2013) nous a permis, en cette première année, d'obtenir des résultats encourageants. Cette activité sera notre priorité pour le prochain quadriennal. La thèse de X. ZHUANG nous permettra d'explorer le potentiel et les limites des dispositifs ME développés dans ce projet.

Dans le cadre de l'étude des capteurs magnétiques, d'autres activités portant sur des dispositifs, tels les flux-gates hybrides, sont poursuivies. Cependant, ces travaux ne sont pas prioritaires et leur mise en oeuvre dépendra des résultats des appels à projet déposés.

Découlant des travaux de thèse M. Timofeeva, la réalisation d'un capteur hybride Champ Electrique - Champ Magnétique est envisagée, l'objectif étant d'atteindre des performances de $1 \text{ pT}/\sqrt{\text{Hz}}$ à 1 Hz et de $1 \mu\text{V}/\text{m}/\sqrt{\text{Hz}}$ en bruit blanc.

AXE 2 : Études et modélisations d'instruments pour des mesures magnétiques De manière plus générale, nos objectifs en magnétométrie restent l'étude, l'optimisation, la réalisation, la mise en oeuvre de divers capteurs magnétiques à haute sensibilité et la réalisation de démonstrateurs mettant en oeuvre un contrôle, une mesure ou l'observation d'un signal magnétique. L'évaluation de leurs performances intrinsèques et extrinsèques doit également permettre de clarifier le réel potentiel de ces différents instruments.

Là encore, nos activités se situent en continuité des travaux réalisés. En contrôle non-destructif par courant de Foucault, notre objectif à moyen terme est la mise en oeuvre d'imagerie magnétique pour des applications en tomographie. Une thèse débutera en Septembre 2010 sur l'étude de la problématique de détection de nanoparticules magnétiques pour des applications biologiques ou en cinétique chimique. De plus, pour compléter nos outils d'analyse, nous continuerons notre étude en microscopie magnétique de manière plus ponctuelle, tant que nous n'aurons pas retrouvé de partenaires sur ce sujet.

AXE 3 : Études et modélisations d'instruments pour des mesures bolométriques Compte-tenu des études déjà menées au laboratoire dans le domaine de la bolométrie, deux perspectives connexes seront privilégiées. D'une part, il s'agira de conclure sur l'apport réel de la thermométrie résistive pour la détection de composés organiques volatiles par la réalisation d'un démonstrateur opérationnel et d'autre part, de développer un bolomètre dédié à des applications de contrôle non destructif thermique. Dans ces deux cas, une approche systémique sera menée.

La quatrième édition de notre école thématique internationale en préparation Enfin, nous travaillons, dès à présent, à l'organisation de la quatrième édition de notre école thématique internationale. Elle devrait avoir lieu au début du prochain quadriennal, en 2012. Cette école apparaît maintenant comme un moment clef dans la communauté, principalement européenne, et devrait permettre d'une part de faire l'état de l'art des développements actuels et d'autre part de répondre aux besoins d'une formation à un haut niveau dans le domaine.

8

Équipe Automatique

8.1	Membres permanents	73
8.1.1	Auto-évaluation	73
8.1.1.1	Opportunités	73
8.1.1.2	Risques	74
8.1.2	Perspectives scientifiques	74
8.1.2.1	Thème 1 : Identification des systèmes	74
8.1.2.2	Thème 2 : Observation des systèmes	74
8.1.2.3	Thème 3 : Commande des systèmes	75
8.1.2.4	Aspects méthodologiques	76
8.2	Nouvelle dynamique de recherche au sein de l'équipe	76

Equipe Automatique

Responsable : Mondher Farza

L'équipe automatique continuera à développer une recherche fondamentale, méthodologique et appliquée sur les problématiques de l'identification, l'observation et la commande des systèmes dynamiques avec des applications privilégiées aux réacteurs (bio)chimiques, systèmes électro-mécaniques et transactions sécurisées. Une attention particulière sera accordée aux développements susceptibles de favoriser des ANR et des collaborations industrielles.

8.1 Membres permanents

L'équipe comprend 15 membres permanents dont 13 membres produisant au sens des critères AERES modulés par les diverses responsabilités assumées par les membres de l'équipe.

Prénom et nom	Fonction	Thèmes	Arrivée
Tarek Ahmed-Ali	PR ENSICAEN	T2 T3	2008
Guy Binet	MC UFR Sciences	T2 T3	
Estelle Cherrier	MC ENSICAEN	T2	2007
Philippe Dorléans	MC UFR Sciences	T2 T3	
Mondher Farza	PR UFR Sciences	T2 T3	
Miloud Frikel	MC ENSICAEN	T1 T2	2008
Olivier Gehan	MC ENSICAEN	T2 T3	
Fouad Giri	PR UFR Sciences	T1 T2 T3	
Mohammed M'Saad	PR ENSICAEN	T1 T2 T3	
Eric Magarotto	MC UFR Sciences	T2 T3	
Jean-François Massieu	MC IUT Caen	T1 T2 T3	
Eric Pigeon	MC UFR Sciences	T1 T2 T3	
Mathieu Pouliquen	MC IUT Caen	T1 T2	
Boubekeur Targui	MC IUT Caen	T1 T2 T3	
Vincent Van Assche	MC UFR Sciences	T1 T3	2009

8.1.1 Auto-évaluation

8.1.1.1 Opportunités

- Le projet de création d'une future INSA de Normandie à partir de la fusion de l'INSA de Rouen et de l'ENSICAEN et la création d'une filière ingénieur au sein de l'UCBN devraient permettre de favoriser l'émergence d'un Master Recherche EEA plus attractif.
- Les nouvelles orientations des pôles de compétitivité TES et MOV'EO semblent être relativement favorables pour les applications de la théorie des systèmes. L'équipe ne ménagera aucun effort pour tirer le meilleur profit de cette nouvelle dynamique afin de développer ses collaborations industrielles.
- La participation de l'équipe au réseau d'excellence HYCON2 va permettre de développer et de conforter ses collaborations nationales et internationales.
- La pyramide des âges de l'équipe nous laisse espérer deux possibilités de recrutement en MC dans les prochaines années. Nous profiterons d'une telle opportunité pour développer et élargir les compétences actuelles dans le cadre des thèmes développés par l'équipe.

8.1.1.2 Risques

- Le vivier national des doctorants en automatique n'est pas assez riche pour pouvoir définir et réaliser sereinement des projets de recherche. Nous comptons sur un vivier relativement riche de doctorants en co-tutelles avec nos collègues des écoles d'ingénieurs en Tunisie et au Maroc.
- Aucun nouveau recrutement n'est prévu au cours du prochain quadriennal pour alléger les charges pédagogiques des membres de l'équipe. Ces charges sont actuellement trop importantes et risquent de s'alourdir avec les nouvelles habilitations de l'UFR de Sciences et de l'ENSICAEN.

8.1.2 Perspectives scientifiques

Les activités de recherche de l'équipe s'inscrivent dans la continuité. Nous attacherons une attention particulière pour que notre production scientifique soit confortée par des publications dans les revues phares de notre communauté et par des projets de recherche attractifs permettant une meilleure visibilité de l'équipe par le monde industriel.

8.1.2.1 Thème 1 : Identification des systèmes

L'identification des systèmes non linéaires a reçu une attention particulière ces dernières années comme l'indiquent les contributions du dernier colloque spécifique à l'identification des systèmes (SYSID'09). L'équipe souhaite conforter ses travaux sur l'identification des systèmes linéaires et contribuer à l'identification des systèmes non linéaires.

- Identification des systèmes par les méthodes des sous-espaces. On exploitera judicieusement le savoir faire de l'équipe en matière d'identification des systèmes et de traitement du signal pour développer une approche expérimentale pour l'identification des systèmes multivariables aussi bien en boucle ouverte qu'en boucle fermée. Cette approche sera utilisée pour traiter les problèmes de géolocalisation, d'identification et d'égalisation des canaux de transmission.
- Identification des systèmes incorporant des blocs non linéaires. Les résultats de l'équipe concernent les modèles de Hammerstein et de Wiener. Le cas de structures plus générales est encore un problème ouvert qui a été à peine abordé par la communauté. On se propose d'étudier les apports des techniques dites évolutionnaires, basées sur des concepts comme la génétique, les essais de particules et le bio-social, pour ce type de problèmes. Le problème d'identification sous-jacent est formulé sous la forme d'un problème d'optimisation de dimension finie en s'appuyant sur une représentation paramétrique. Des résultats préliminaires ont d'ores et déjà été obtenus.
- Identification des systèmes non linéaires. Les résultats de l'équipe sur la synthèse des observateurs adaptatifs des systèmes non linéaires par rapport aux variables d'état et aux paramètres seront particulièrement utilisés pour initier une étude sur l'identification des systèmes non linéaires. Des résultats préliminaires sur l'identification des retards ont déjà été obtenus.

8.1.2.2 Thème 2 : Observation des systèmes

Les contributions de l'équipe en matière d'observation des systèmes non linéaires offrent de nombreuses perspectives d'investigation relatives aux problèmes ouverts suivants.

- Caractérisation des systèmes multivariables uniformément observables. Cette étude consiste à exhiber une forme canonique caractérisant la classe la plus large des systèmes uniformément observables. Des résultats préliminaires ont d'ores et déjà été obtenus et nous travaillons actuellement sur leur généralisation.
- Synthèse d'observateurs pour des systèmes non linéaires en présence de bruits de mesure. Cette étude est principalement motivée par la sensibilité des observateurs du type grand gain aux bruits de mesure. Pour pallier ce problème, l'approche de synthèse d'observateur à grand gain sera reconsidérée. Contrairement à l'observateur classique dont le gain se règle à travers un paramètre constant

(pris généralement assez grand), le réglage du nouvel observateur se fait à travers un paramètre variant dans le temps et régi par une équation différentielle de type Riccati qui prend en compte la variance du bruit de la sortie. La classe de systèmes considérés est caractérisée par une forme canonique uniformément observable multi-sorties formée par plusieurs blocs avec un couplage non triangulaire.

- Synthèse d'observateurs pour les systèmes à retard. Cette étude concerne plusieurs problèmes ouverts inhérents aux systèmes à retard. Un des problèmes est relatif à la classe des systèmes exhibant des retards inconnus qui interviennent au niveau des états non mesurés. Il s'agit de caractériser la classe des systèmes pour lesquels l'identification des retards est possible et de procéder par la suite à la synthèse d'observateurs pour l'estimation conjointe des états et des retards. Un autre problème d'actualité concerne les systèmes avec des sorties retardées. Nous avons déjà proposé des solutions lorsque le retard est constant en considérant aussi bien le cas d'un retard connu qu'inconnu. Le problème du retard variable, même lorsque la variation de ce retard est connue, reste ouvert et très peu de travaux ont abordé ce sujet.
- Synthèse d'observateurs à entrées inconnues. On étendra d'abord nos travaux à une large classe de systèmes multivariés uniformément observables non linéaires par rapport aux entrées inconnues. Des conditions suffisantes pour l'estimation conjointe des variables d'état et des entrées inconnues seront établies. Nous exploiterons ensuite ces résultats pour l'observation des systèmes singuliers. Les observateurs ainsi conçus seront judicieusement utilisés pour traiter des problèmes de diagnostic et de supervision.
- Observation des systèmes avec contraintes de communication. Cette étude concerne le cas des systèmes continus où les mesures ne sont disponibles qu'aux instants d'échantillonnage. De plus, ces mesures ne sont pas nécessairement précises et sont généralement entachées d'erreurs de quantification ou de transmission à travers un réseau. On considérera des observateurs continus spécifiques pour ces systèmes et on déterminera pour chacun de ces observateurs des conditions suffisantes, sur les gains, les périodes d'échantillonnage et les périodes de blocage des mesures, pour préserver la stabilité et assurer la convergence asymptotique de versions continues discrètes de ces observateurs. Pour ce faire, on exploitera les concepts de la théorie de stabilité des systèmes hybrides et plus particulièrement l'approche de Lyapunov pour des systèmes dits impulsifs.

8.1.2.3 Thème 3 : Commande des systèmes

Les contributions de l'équipe sur la commande des systèmes non linéaires ont été appréciées par notre communauté comme en témoignent les nombreuses communications en la matière. Outre leur valorisation par des publications dans des revues, ces contributions seront confortées et développées à travers les problématiques suivantes.

- Commande des systèmes multivariés avec retour de sortie. Cette étude consiste à étendre le contexte de synthèse de systèmes de commande du type grand gain développé au sein de l'équipe à des classes de systèmes commandables et uniformément observables. Ces systèmes sont caractérisés par une forme canonique composée par une cascade de sous-systèmes avec des couplages non triangulaires. Pour ce faire, on exploitera judicieusement le concept de dualité observation-commande pour tirer le meilleur profit de nos résultats sur l'observation des systèmes.
- Commande adaptative non linéaire. On exploitera d'abord nos résultats sur l'observation et la commande pour proposer des algorithmes de commande adaptative en supposant que la condition d'excitation persistante requise pour la convergence paramétrique est satisfaite. On cherchera ensuite à relâcher la condition d'excitation persistante en utilisant des modifications appropriées des lois d'adaptation paramétriques.
- Commande des systèmes échantillonnés. Cette étude consistera à proposer des lois de commande échantillonnées à partir des classes de systèmes de commande continus disponibles, notamment celles proposées par l'équipe. La classe de systèmes considérée est constituée par un ensemble de

systèmes incertains en cascade. On montrera que la loi de commande échantillonnée peut être obtenue en ajoutant à la loi de commande continue des termes dépendant de la période d'échantillonnage. Cette modification permet de préserver la stabilité du système de commande échantillonné tout en améliorant les performances par rapport au cas d'une simple émulation du système de commande continu.

- Commande des systèmes linéaires en présence de contraintes sur l'entrée et/ou la sortie. On traitera plus particulièrement le cas des systèmes continus avec retard en présence d'actionneurs et/ou de capteurs saturants. Une attention particulière sera réservée au problème de poursuite et donc à la caractérisation des séquences de références compatibles avec les contraintes considérées.

8.1.2.4 Aspects méthodologiques

Un effort important sera consacré au sein de l'équipe à la recherche méthodologique pour développer des collaborations universitaires et industrielles. Ces applications sont particulièrement motivées par notre souci de participer activement à des projets ANR avec nos collègues.

- Une approche expérimentale pour la conduite des bioréacteurs est en cours de développement en collaboration avec le "Center of Biological Engineering" de l'Université de Minho de Braga au Portugal. Des résultats préliminaires ont été obtenus avec l'approche du type grand gain développée par l'équipe et une validation expérimentale est prévue au cours de l'année 2011.
- Commande des systèmes électro-mécaniques. Les connaissances acquises sur les aspects de modélisation, d'observation et de commande des convertisseurs statiques et des moteurs asynchrones seront judicieusement exploités pour développer une solution potentielle au problème de commande des machines électriques. Cette solution pourrait être valorisée à travers des collaborations industrielles. Des résultats probants ont été obtenus suite à des validations expérimentales d'un système de commande sans capteur de vitesse.
- Développement d'un système de commande d'accumulateurs électriques embarqués en collaboration avec le LUSAC de Cherbourg. Ce travail, qui s'inscrit dans le contexte du développement durable, est particulièrement motivé par notre souci de participer activement à des projets ANR. Sa partie fondamentale sera développée au sein de l'équipe.
- Développement de méthodologies pour l'identification et l'égalisation des canaux de transmission CDMA. Une réflexion est faite au sein de l'équipe sur la conception d'une plate forme expérimentale pour valider nos résultats fondamentaux sur la transmission sécurisée.

8.2 Nouvelle dynamique de recherche au sein de l'équipe

L'équipe automatique a souffert de l'absence d'une conception commune de la recherche comme le montre la répartition de sa production scientifique par rapport à ses membres. On distingue trois groupes comme l'indique clairement les fiches individuelles des membres de l'équipe.

- Des membres publiants au sens de l'AERES qui ont eu l'opportunité de travailler dans des environnements scientifiques prestigieux, notamment le LAG, LAGEP, LSS.
- Des membres peu publiants au sens de l'AERES qui se sont bien investis dans des responsabilités pédagogiques et/ou collectives au détriment de la recherche. Nous les avons considérés comme produisant compte tenu de leurs responsabilités et du gradient prometteur de leurs activités de recherche.
- Deux membres non publiants qui ont toutes les connaissances requises pour rejoindre le groupe des publiants. L'un effectue une bonne recherche sur le traitement du signal et ses applications avec un investissement important sur des réalisations pédagogiques. Il participe régulièrement à l'encadrement des stages et apporte une aide potentielle aux nouveaux arrivants en enseignement. L'autre s'est complètement investi dans des activités pédagogiques et l'organisation des séminaires.

de l'équipe. Le nécessaire a été fait pour qu'il puisse retrouver le chemin de la recherche.

Au cours du dernier quadriennal, l'équipe automatique a initié une dynamique interne favorisant l'émergence de nouveaux projets de recherche pour ses membres peu producteurs au sens de l'AERES. Un travail important au sein de l'équipe a été consacré durant le dernier quadriennal sur l'identification des principales difficultés de nos collègues à faire de la recherche. On distingue

- Un investissement important pour la rénovation de nos enseignements afin de rendre plus attractives nos filières. Cette rénovation pédagogique a mobilisé beaucoup d'énergie ces dernières années : il fallait conforter ses bases fondamentales en théorie des systèmes et acquérir une bonne expérience sur les aspects de mise en oeuvre sous-jacents. Ceci a permis de concevoir des systèmes d'apprentissage dédiés et des manipulations bien appropriées pour la validation des approches expérimentales développées. Beaucoup de nos étudiants utilisent avec succès ces approches lors de leurs stages dans l'industrie. Des membres de notre équipe ont reçu un prix pour une contribution pédagogique présentée au Colloque sur l'Enseignement des Technologies et des Sciences de l'Information et des Systèmes organisé à Bordeaux.
- Un investissement important sur les systèmes non linéaires pour pouvoir faire des contributions fondamentales et méthodologiques en la matière. En effet, la plupart de nos membres avaient fait leur recherche sur les systèmes linéaires.
- Une implication très réduite dans les comités d'encadrement des thèses et des stages de master recherche.
- Des responsabilités collectives importantes, notamment la responsabilité de plusieurs diplômes au niveau LMD option EEA à l'UFR des Sciences ainsi que de deux licences professionnelles.

A partir de ce constat nous avons décidé de réduire les responsabilités collectives et tirer le meilleur profit du potentiel pédagogique que nous avons acquis pour consacrer plus de temps pour la recherche. Un projet de recherche a été défini avec chaque membre peu produisant de l'équipe à partir de ses connaissances actuelles des thèmes de l'équipe. Des groupes de travail ont été constitués à partir des projets de recherche, notamment

- Le développement de méthodologies pour la conduite des réacteurs (bio)chimiques à partir des résultats fondamentaux de l'équipe en matière d'observation et de commande des systèmes non linéaires et du savoir faire en matière d'identification et de commande des systèmes linéaires. Une attention particulière est réservée au problème d'estimation des vitesses de réactions ou de certains paramètres cinétiques. Des validations expérimentales sont prévues chez nos collègues du LAAS et du Centre de Biotechnologie de l'Université de Minho de Braga au Portugal.
- Le développement de méthodologies de commande des systèmes électro-mécaniques et leur validation sur les plates-formes expérimentales disponibles au laboratoire, en l'occurrence le banc moteur asynchrone et le pendule inverse. Cette recherche est à l'ordre du jour dans notre communauté comme en témoignent les activités du groupe de travail Commande des Entraînements Electriques (CE2) du GDR MACS. Une attention particulière est réservée au sein de notre équipe à la caractérisation des sous-espaces inobservables et à la synthèse des observateurs pour cette classe de systèmes.
- Le développement de méthodologies pour l'identification et l'égalisation des canaux de transmission CDMA avec une ouverture sur les problématiques de transactions sécurisées et de téléopération.
- L'étude des problèmes d'échantillonnage lors de la mise en oeuvre des observateurs et des systèmes de commande non linéaires développés pour les systèmes continus.

Ces projets vont permettre à nos collègues de recouvrer une dynamique de recherche fructueuse. Des résultats prometteurs ont été obtenus et présentés dans des communications internationales et d'autres

sont en cours de rédaction pour des revues internationales.

Pour conforter cette dynamique de recherche, nous avons prévu d'organiser régulièrement des séminaires (un à deux par mois) et de réaliser une auto-évaluation annuelle en présence d'un ou deux collègues de notre communauté que nous inviterons à l'occasion. Ces séminaires seront animés en fonction de l'avancement de nos projets de recherche et porteront sur les aspects suivants.

- des enseignements du type master recherche pour les bases fondamentales sur les systèmes non linéaires,
- des tours d'horizon guidés sur les récents développements en relation avec nos thèmes de recherche,
- des présentations de résultats probants sur nos thèmes de recherche développés par des collègues de notre communauté, en l'occurrence ceux que nous inviterons lors de notre auto-évaluation,
- des présentations de nos doctorants sur l'avancement de leurs travaux, les comités d'encadrement en profiteront pour en faire le point.

Par ailleurs, chaque membre est fortement incité à participer aux groupes de travail des GDR MACS et ISIS. C'est une opportunité de se positionner sur le plan national et de pouvoir engager des collaborations.

