RAPPORT DE CONJONCTURE

DU COMITÉ NATIONAL DE LA RECHERCHE SCIENTIFIQUE

ÉDITION 2014

Extrait

CNRS ÉDITIONS

15, rue Malebranche – 75005 Paris

ISBN : 978-2-271-08746-1

SECTION 06

SCIENCES DE L'INFORMATION : FONDEMENTS DE L'INFORMATIQUE, CALCULS, ALGORITHMES, REPRÉSENTATIONS, EXPLOITATIONS

Extrait de la déclaration adoptée par le Comité national de la recherche scientifique réuni en session plénière extraordinaire le 11 juin 2014

La recherche est indispensable au développement des connaissances, au dynamisme économique ainsi qu'à l'entretien de l'esprit critique et démocratique. La pérennité des emplois scientifiques est indispensable à la liberté et la fécondité de la recherche. Le Comité national de la recherche scientifique rassemble tous les personnels de la recherche publique (chercheurs, enseignants-chercheurs, ingénieurs et techniciens). Ses membres, réunis en session plénière extraordinaire, demandent de toute urgence un plan pluriannuel ambitieux pour l'emploi scientifique. Ils affirment que la réduction continue de l'emploi scientifique est le résultat de choix politiques et non une conséquence de la conjoncture économique.

L'emploi scientifique est l'investissement d'avenir par excellence

Conserver en l'état le budget de l'enseignement supérieur et de la recherche revient à prolonger son déclin. Stabiliser les effectifs ne suffirait pas non plus à redynamiser la recherche : il faut envoyer un signe fort aux jeunes qui intègrent aujourd'hui l'enseignement supérieur en leur donnant les moyens et l'envie de faire de la recherche. On ne peut pas sacrifier les milliers de jeunes sans statut qui font la recherche d'aujourd'hui. Il faut de toute urgence résorber la précarité. Cela suppose la création, sur plusieurs années, de plusieurs milliers de postes supplémentaires dans le service public ainsi qu'une vraie politique d'incitation à l'emploi des docteurs dans le secteur privé, notamment industriel.

Composition de la section

Frédérique BASSINO (présidente de section) ; Hugo GIMBERT (secrétaire scientifique) ; Leila AMGOUD ; Guillaume BESLON ; Patricia BOUYER-DECITRE ; Giuseppe CASTAGNA ; Rachida CHABANE ; Marie-Pierre COMBEAU ; Évelyne CONTEJEAN ; Marie-Odile CORDIER ; Pierre CRESCENZO ; Pascal DAYRE ; Frédéric GARDI ; Isabelle GUERIN-LASSOUS ; Mathieu LATAPY ; Pierre LOPEZ ; Frédéric MAGNIEZ ; Yann PONTY ; Guillaume RASCHIA ; Marie-Christine ROUSSET ; Sophie TISON.

Résumé

Rédigé du printemps à l'automne 2014, ce rapport fait un état de l'art de la recherche des 4 dernières années en informatique sur les thèmes des sciences de l'information relevant de la section 6 : fondements de l'informatique, calculs, algorithmes, représentations, exploitations.

L'informatique est une science d'une dualité historique, mêlant études de grandes questions fondamentales, et recherche de solutions, éventuellement logicielles, à des problèmes concrets. Cette période s'achève ainsi par la deuxième médaille d'or attribuée à un informaticien par le CNRS pour des travaux mêlant théorie et pratique. Loin d'être isolée, cette reconnaissance du CNRS s'accompagne de plusieurs autres grandes réussites françaises récompensées par les prix les plus prestigieux. Après un rapide tour d'horizon du paysage français de la communauté informatique, et des thèmes émergents qui ont influencé la discipline, ce rapport présentera un état de la recherche informatique dans le monde selon 6 thèmes : (1) Calcul : algorithmes, combinatoire, protection de l'information ; (2) Fondements de l'informatique, preuve et vérification ; (3) Programmation et architecture logicielle, systèmes et réseaux ; (4) Données et connaissances – Intelligence artificielle et interactions ; (5) Aide à la décision et recherche opérationnelle ; (6) Nouvelles interactions avec les autres sciences. Seront mis en avant, pour chaque thème, les nouvelles directions et les nouveaux paradigmes, des résultats récents parmi les plus marquants, ainsi que le positionnement de la recherche faite en France.

Introduction

Alors que Gérard Berry, titulaire de la première chaire du Collège de France consacrée à l'informatique, vient de recevoir la médaille d'or du CNRS, notamment pour le langage Esterel dont il est l'inventeur et qui est largement utilisé dans l'industrie, ainsi que pour ses travaux plus récents sur la programmation diffuse, c'est-à-dire à destination d'objets connectés, la communauté française relevant des thématiques de la section 6 peut se féliciter de l'impact national et international de son activité scientifique. Ainsi, 11 bourses de l'European Research Council (ERC) ont été obtenues depuis 2010 et 3 prix internationaux prestigieux durant la seule année 2013 :

• Le prix John von Neumann a été décerné par l'Institute for Operations Research and the Management Sciences (INFORMS) à Michel Balinski pour l'ensemble de ses contributions en recherche opérationnelle et aide à la décision, qui ont marqué les domaines de la combinatoire polyédrale, de l'optimisation combinatoire et de la programmation linéaire en nombres entiers, mais aussi de la théorie du choix social et de la théorie du vote. Ses travaux les plus récents, sur le vote par jugement majoritaire, ont reçu un écho médiatique important ; ainsi l'un de ses systèmes de vote est effectivement utilisé à Zurich.

• Le prix Gödel co-décerné par l'European Association for Theoretical Computer Science et le pôle algorithmique et informatique théorique de l'Association for Computing Machinery (ACM) à Antoine Joux et deux autres scientifiques pour leurs travaux portant sur la notion de couplage en cryptographie, une technique jusque-là réservée au domaine des mathématiques pures. En rendant ce procédé utilisable en cryptographie, ils ont permis à des nouvelles fonctionnalités de voir le jour, notamment dans le contexte de la lutte contre le piratage des contenus multimédia.

• Le Software System Award décerné par l'ACM au logiciel Coq, récompensant une équipe de 9 principaux contributeurs dont 7 sont actuellement en poste en France. Coq obtient ainsi la reconnaissance d'être devenu un assistant de preuve précieux pour l'écriture et la vérification de programmes, mais aussi pour résoudre des conjectures mathématiques hors de portée de preuves non automatisées.

Le paysage français de la recherche en informatique est particulier avec plus de 3 000 enseignants-chercheurs en section CNU 27 soit 10 fois plus que de chercheurs CNRS ou encore de chercheurs INRIA relevant de cette discipline. D'autres établissements, tels le CEA, comptent également des chercheurs en informatique toutefois en plus petits nombres. L'INS2I est structuré en seulement 47 laboratoires CNRS (propres ou mixtes) dont environ la moitié relèvent principalement de la section 6, et parmi ceux-là une dizaine comptent plus de 100 permanents. Un réseau des Groupes de Recherche du CNRS structure thématiquement cet ensemble de chercheurs et assure l'animation des différents domaines de recherche. Ces GdR regroupent chacun plusieurs centaines de membres permanents, et autant de doctorants et postdocs, 6 d'entre eux sont principalement rattachés à la section 6.

La proportion de femmes dans la profession demeure faible. Elles représentent moins de 20 % des chercheur-e-s de la section 6, et à peine 25 % des enseignant-e-s-chercheur-e-s de la section CNU 27.

Le nombre de doctorants dans le domaine est important. De l'ordre de 500 d'entre eux soutiennent chaque année leur thèse et trouvent un emploi dans la recherche académique ou dans le secteur privé, leur doctorat n'étant dans ce dernier cas pas reconnu la plupart du temps.

La recherche en sciences de l'information est traversée par une double influence : d'une part l'étude de grandes questions fondamentales qui ont des répercussions économiques et sociétales tout en contribuant à l'évolution même de la discipline, et d'autre part la recherche de solutions concrètes, éventuellement logicielles, à des problèmes pratiques.

Ainsi P vs NP est l'un des 7 problèmes du millénaire selon le Clay Mathematics Institute, la fiabilité de nombreux systèmes cryptographiques repose sur la complexité de problèmes arithmétiques tels que celui du logarithme discret sur les courbes elliptiques. Enfin le développement de l'informatique quantique remet en question la nature du calcul telle que conçue par Turing, et est déjà en partie utilisée dans certaines applications.

L'évolution du monde numérique alimente aussi la recherche en informatique, de ses applications jusqu'à ses travaux fondamentaux. Ces dernières années ont ainsi vu émerger de nouvelles problématiques qui concernent tous les axes de recherche de la section 6 et sont au cœur des évolutions récentes de la recherche en informatique :

- le traitement à la volée ou a posteriori des flux de données temporels, dynamiques, et hétérogènes ;
- les problèmes de sécurité : sûreté de fonctionnement, protection de la vie privée, cryptographie ;
- la gestion des ressources en programmation et la certification des programmes, le développement de logiciels libres ;
- l'étude de nouveaux modèles ou paradigmes de calcul et le calcul haute performance ;
- les réseaux de communication : augmentation des connexions et des interactions, des volumes de données échangées ou partagées.

Ce rapport de conjoncture a été préparé selon un découpage par thèmes du périmètre scientifique de la section. Dans chaque thème, la recherche menée ainsi qu'une sélection des nouvelles directions et nouveaux paradigmes de la discipline sont présentés. Vient ensuite la présentation de contributions parmi les plus marquantes, tant au niveau national qu'international, afin de mettre en évidence le positionnement de la recherche faite en France. Compte tenu des contraintes de place, des choix ont été effectués, afin de ne conserver que quelques-unes des avancées et directions de recherche les plus significatives.

I. Calcul : algorithmes, combinatoire, protection de l'information

Un volet important du traitement de l'information s'articule autour de la notion de calcul avec tout ce qu'elle comporte : algorithmes, arithmétique des ordinateurs, calcul formel, protection de l'information, ainsi que l'analyse des structures associées qui sont le propre de la combinatoire, de la théorie des graphes ou encore des systèmes dynamiques. Ce domaine de recherche constitue une interface très active entre mathématiques discrètes et informatique.

A. Algorithmes : méthodes génériques, analyse de complexité

La recherche en algorithmique s'organise principalement autour du développement de techniques permettant de rapprocher bornes inférieures (résultats d'impossibilité) et supérieures (algorithmes) tant sur la complexité de problèmes informatiques que sur la qualité des solutions produites. L'explosion des données massives a suscité un intérêt croissant pour les modèles de calcul qui prennent en compte un accès contraint aux données, comme le calcul distribué, en ligne, de streaming, ou encore les algorithmes sous-linéaires. L'exportation de la notion d'algorithme à d'autres sciences s'est traduite par l'étude algorithmique des systèmes dits naturels engendrés par les réseaux sociaux ou encore en biologie. La recherche en France dans ce domaine est d'un très bon niveau international.

Plusieurs rapprochements thématiques ont eu lieu, en particulier avec les mathématiques. Un exemple typique est celui de la recherche opérationnelle, ou plus précisément de la programmation mathématique. Un des résultats les plus marquants concerne les limites de la programmation linéaire comme outil pour concevoir des algorithmes d'approximation, par rapport à ceux plus complexes de la programmation semi-définie. Une séparation vient d'être établie sur un problème pratique important, le voyageur de commerce, résolvant ainsi une question formulée il y a 20 ans par Yannakakis. La beauté de ce résultat réside dans le lien nouveau qu'il établit entre programmes linéaires et complexité de la communication quantique. Toute une série de résultats similaires tant théoriques que pratiques ont été établis à sa suite. Les chercheurs européens, dont quelques uns en France, ont joué un rôle moteur dans l'établissement de ces ponts.

Parallèlement à l'analyse de la complexité des problèmes, l'analyse (en moyenne) des algorithmes eux-mêmes a remporté quelques succès en permettant de comprendre et prédire le comportement d'algorithmes plus complexes exécutés sur des instances plus réalistes. Ces avancées ont été réalisées grâce à l'apport de méthodes génériques issues de la combinatoire analytique, fructueux mariage de la combinatoire formelle et de l'analyse complexe, domaine qui s'est développé sous l'impulsion de Philippe Flajolet et Robert Sedgewick.

B. Théorie des graphes, algorithmique des graphes

Les activités de recherche autour des graphes portent à la fois sur des aspects structurels et algorithmiques. La communauté française nombreuse dans ce domaine est internationalement reconnue.

Parmi les principales avancées, on peut citer le développement de toute une théorie pour traiter les graphes en les voyant comme des objets limites. La théorie est bien comprise et unifiée pour les graphes denses et a de nombreuses ramifications vers des sujets variés comme les algorithmes sous-linéaires et l'échantillonnage. De nouveaux progrès ont permis de résoudre d'importantes conjectures relatives à certaines partitions de graphes. Fort

de la compréhension des graphes denses, un recentrage s'est effectué vers l'étude des graphes creux ou épars, véritable enjeu pour la modélisation des grands graphes issus des applications récentes comme les réseaux au sens large. Plusieurs résultats obtenus en France ont permis de développer un cadre unifié pour traiter les graphes creux.

Un résultat notable à l'interface des mathématiques discrètes et de l'algorithmique concerne la version constructive du lemme local de Lovasz. Ce lemme est une variante de la méthode probabiliste qui permet de construire des objets à partir d'évènements dont chacun ne dépend que d'un petit nombre d'autres évènements. Les applications sont nombreuses et vont de la théorie des graphes extrémaux aux problèmes de routage et de coloration. La construction algorithmique proposée par Moser a ouvert de nouveaux champs d'applications qui ont été investis en partie par la communauté française.

Dans le domaine de la complexité paramétrée, les progrès récents les plus significatifs sont des résultats montrant l'existence de noyaux polynomiaux pour de larges familles de problèmes paramétrés ainsi que le développement de techniques de bornes inférieures sur la taille des noyaux. Dans ce contexte, l'optimisation des complexités existantes, les liens entre les méthodes paramétrées et les méthodes d'approximations sont des enjeux importants.

C. Combinatoire, systèmes dynamiques discrets, algorithmique du texte

À l'interface entre informatique théorique, mathématiques discrètes, théorie des probabilités et physique théorique, une large communauté s'intéresse aux structures discrètes aléatoires ou algébriques issues de ces disciplines. Cette interaction est particulièrement riche en France et constitue une des forces de la combinatoire. Des contributions notables ont ainsi été obtenues à l'interface avec la physique (physique statistique et renormalisation) et dans des interactions renouvelées avec des domaines des mathématiques centrés sur le calcul, comme le calcul moulien ou le calcul opéradique. Parmi les nombreux résultats marquants récemment obtenus citons l'élaboration d'interprétations combinatoires d'objets extrêmement difficiles à calculer, notamment autour des polynômes de MacDonald, la résolution d'un problème ouvert depuis plus de 20 ans sur les systèmes de particules en interaction (modèles ASEP et TASEP) ou encore la preuve de la conjecture de Razumov-Stroganov intimement liée aux travaux sur les matrices à signes alternants.

L'étude des systèmes dynamiques discrets s'est principalement développée récemment autour des domaines de la dynamique symbolique, de la théorie des pavages (et de leurs espaces), et des automates cellulaires. La notion de substitution agissant sur des mots ou sur des tuiles est un concept crucial dans ce contexte. Les substitutions permettent en effet d'exprimer la dynamique de la renormalisation, de l'induction (l'étude des échanges d'intervalles en est une bonne illustration), mais aussi des notions issues de la calculabilité : elles sont utilisées pour coder et construire des zones de calcul. Dans le cadre de la dynamique multidimensionnelle, la communauté française a ainsi contribué à montrer que les outils de complexité et de calculabilité sont essentiels pour la description et la compréhension des systèmes dynamiques symboliques, en obtenant en particulier la caractérisation des sous-shifts effectifs de dimension d comme projection de sous-shifts de type fini de dimension (d+1).

L'algorithmique du texte a connu un essor important en liaison en particulier avec les travaux sur des problèmes issus du traitement des séquences génomiques (représentation compressée des graphes de De Bruijn grâce aux filtres de Bloom, table des suffixes dynamique et transformée de Burrows-Wheeler, toutes trois utilisées au sein des algorithmes d'assemblage pour accélérer et limiter la consommation

mémoire du traitement des données produites par séquençage haut-débit). La France est ainsi très reconnue dans les domaines fondamentaux, mais elle souffre d'un manque de valorisation, la plupart des outils utilisés dans les laboratoires de biologie n'étant pas des logiciels français. Il existe cependant quelques exceptions notables, par exemple, pour la recherche des duplications en tandem (logiciels PhyML et STAR) et le repliement comparatif des ARN (CARNAC), où des outils constituant l'état de l'art du domaine sont développés en France.

D. Calcul arithmétique et formel, codage et cryptographie

L'arithmétique des ordinateurs et le calcul formel s'intéressent à la manipulation des objets fondamentaux de l'arithmétique et de l'algèbre (nombres, polynômes, séries, solutions d'équations différentielles...). Lorsque l'on conçoit les briques de base du calcul, sur lesquelles repose tout l'édifice, l'efficacité et la sécurité sont alors primordiales. La cryptographie, qui s'est longtemps appuyée sur des objets algébriques relativement simples (calcul modulaire), fait maintenant davantage appel à des objets nettement plus complexes relevant du calcul formel, comme les courbes elliptiques et les réseaux euclidiens.

En cryptographie, les thèmes et résultats qui ont émergé ces dernières années concernent la sécurité pour le cloud computing avec notamment les chiffrements fonctionnels (qui permettent plusieurs niveaux de déchiffrement en fonction de la clé secrète détenue) et les chiffrements complètement homomorphes (qui permettent de déléguer sur un serveur un calcul sur des données qui restent chiffrées tout au long du calcul), ainsi que les attaques physiques par canaux cachés et la résistance aux fuites partielles d'information. Enfin, les thématiques autour de la protection de la vie privée et de l'anonymat ont pris une importance considérable.

Sur des aspects plus algorithmiques, les avancées majeures concernent la sécurité de systèmes de chiffrement qui repose maintenant sur des complexités en pire cas, et non plus en moyenne (en utilisant des problèmes de réseaux euclidiens), en particulier la découverte d'un algorithme efficace pour le calcul du logarithme discret dans les corps finis de petite caractéristique.

En arithmétique des ordinateurs, la reproductibilité numérique prend de l'importance pour les applications de calcul intensif. La conception automatique de bibliothèques de fonctions élémentaires ou spéciales, afin de s'adapter très vite à un nouveau contexte (format cible, processeur utilisé, exigences en matière de vitesse, précision, etc.) est au cœur des travaux actuels, pour lesquels la communauté française est bien positionnée. Les thèmes traités vont de l'arithmétique matérielle et logicielle pour la cryptographie à la génération de bibliothèques et d'opérateurs d'arithmétiques flottantes ou fixes, en passant par la validation et la preuve automatique. On peut néanmoins regretter la quasi-absence d'une industrie française des semi-conducteurs à l'exception notable de STMicroelectronics.

II. Fondements de l'informatique, preuve et vérification

L'informatique fondamentale s'attache à étudier et à systématiser les démarches, parfois empiriques, mises en œuvre dans le développement logiciel, ainsi qu'à concevoir de nouveaux cadres ou méthodologies. Cette thématique se situe à la croisée d'une variété de domaines tels que la logique, la théorie des automates, les systèmes de réécriture, les algèbres de calcul et encore des systèmes de contraintes. Les applications sont diverses, allant de la conception à l'analyse de codes sûrs (codes embarqués, systèmes répartis ou mobiles, protocoles de sécu-

rité, ou plus généralement, tous les codes dont le bon fonctionnement est crucial d'un point de vue économique, médical ou sociétal). L'implication de la communauté française dans ces domaines de recherche est très importante avec de nombreux projets et réseaux européens.

A. Avancées majeures

Une avancée récente importante se situe au cœur de nombreux problèmes de vérification et concerne les réseaux de Petri, des objets introduits dans les années 60 pour modéliser des processus concurrents. Il s'agit du problème crucial de l'accessibilité dans ces réseaux, pour lequel le premier algorithme, très complexe tant d'un point de vue conceptuel que d'un point de vue calculatoire, n'a été proposé qu'au début des années 1980. Après de nombreuses recherches pour développer des solutions plus simples, la France a récemment eu un rôle moteur et décisif, en développant un nouvel algorithme qui repose sur une toute nouvelle approche conceptuellement simple, à savoir le calcul d'invariants inductifs presque semi-linéaires permettant de valider soit l'accessibilité, soit la non-accessibilité. Ces résultats ouvrent de nouvelles perspectives, qui permettront, peut-être dans un futur proche, de mieux cerner la complexité théorique intrinsèque du problème de l'accessibilité dans les réseaux de Petri.

Une autre avancée notable porte sur le format standard XML (eXtensible Markup Language) pour l'échange de données semi-structurées qui est utilisé dans des services web, les bases de données, et pour échanger des données entre applications. La recherche française a attaqué les problématiques liées à l'utilisation de ce format avec des techniques très variées. Elle a ainsi fait des progrès importants en langages de programmation grâce à la définition de systèmes polymorphes pour le traitement de données en format XML ou en analyse statique des transformations de documents en format XML, notamment grâce à l'utilisation de solveurs pour des logiques modales.

Une autre avancée marquante concerne un problème classique en théorie des langages formels : décider si le langage des mots acceptés par un automate peut aussi être exprimé par une formule logique. On sait par exemple que la logique monadique du second ordre a le même pouvoir expressif que les automates finis. Une avancée marquante vient d'être obtenue : la décidabilité de l'appartenance d'un langage aux classes $B\Sigma 2$ et $\Sigma 3$. La résolution de ce problème ouvert depuis quarante ans vient renouveler le sujet de manière profonde grâce à une innovation conceptuelle importante.

Enfin, la dernière avancée majeure retenue ici concerne la théorie homotopique des types. La théorie des types intuitionniste est une formalisation de la logique intuitionniste développée par Martin-Löf dans les années 70 qui se caractérise par la présence de types paramétrés. Un des défis majeurs a été la compréhension du statut de l'égalité. Une nouvelle approche vient de changer notre perception de cette problématique. Voevodsky a réalisé que l'algèbre homotopique fournissait des modèles très riches de la théorie des types, où, intuitivement, l'égalité devient une déformation (une homotopie) permettant de transformer un objet en un autre, généralisant considérablement les idées antérieures. Cela a amené Voevodsky à ajouter un axiome à la théorie des types : l'axiome d'univalence exprime une forme de complétude du type identité vis-à-vis de ces déformations. Il a ainsi établi un pont très surprenant entre logique d'une part et topologie algébrique d'autre part, riche de promesses tant pour la logique et l'informatique que pour les mathématiques. Cette théorie a déjà une influence certaine sur le développement des assistants de preuves comme Coq.

B. Nouvelles directions

Plusieurs développements récents traduisent un basculement vers de nouveaux points de vue ou paradigmes. Ainsi, en vérification, les spécifications et les techniques d'ana-

lyse ne sont plus seulement qualitatives mais également quantitatives : il ne suffit plus seulement de savoir si un système peut atteindre un état interdit, mais avec quelle probabilité cela peut avoir lieu, ou dans quel contexte (bande passante allouée, contrainte d'énergie, etc.), ou encore d'étudier la robustesse des propriétés de systèmes soumis à des perturbations. Dans une autre direction, des liens remarquables ont été établis entre la théorie des jeux et les techniques de vérification ou de contrôle permettant ainsi de modéliser et de synthétiser des protocoles mettant en jeux de multiples agents. L'interaction d'un système avec un ou plusieurs acteurs de l'environnement est alors vue comme un jeu à plusieurs joueurs, et les spécifications comme des objectifs de gain.

Dans le domaine de la sécurité, la communauté est passée de l'analyse de propriétés de sécurité classiques (authentification, confidentialité), basées sur de l'accessibilité, à des propriétés ayant trait à la vie privée (secret du vote, anonymat, non traçabilité, etc.). Un pont important pour les applications a également été établi vers les modèles plus fins de sécurité utilisés en cryptographie.

Enfin, dans le domaine des bases de données, les points de vue ont fortement évolué, en phase avec le développement de nouveaux formats dont la structuration relationnelle est faible (NoSQL) et l'explosion des bases de données massives, hétérogènes, incomplètes, incertaines, et dynamiques. Ainsi, des logiques et automates ont été développés pour capturer ces évolutions. Les nouvelles problématiques posées par l'interrogation efficace de ces bases de données et l'évolution de leurs formats sont nombreuses.

C. Outils développés en France et impact sociétal

La recherche française se distingue ici par le développement de nombreux prototypes, aboutissant pour certains à des outils à fort impact. En sécurité, différents prototypes d'analyse automatique de protocoles ont vu le jour, et trois des quatre principaux outils ont été mis au point en France (ProVerif, APTE, Akiss). En vérification et en analyse statique, une plate-forme a récemment vu le jour (Cosyverif), visant à proposer un format unique de manipulation de systèmes. L'outil d'analyse statique de code Astrée est par ailleurs maintenant commercialisé et régulièrement utilisé par l'industrie.

Enfin, après plus de 20 ans de développement, l'outil de preuve formelle Coq (Calculus of Constructions) a atteint la pleine reconnaissance en recevant en 2014 le prix ACM Software System Award, le plaçant ainsi au même niveau que des standard tels que TeX, TCP/IP, WWW, Apache, Java, Unix et Postscript. Les systèmes de vérification de preuves formelles sont apparus à la fin des années 1960. En plus de la vérification, ils aident l'utilisateur à construire interactivement des preuves à l'aide de stratégies. Ainsi, l'assistant de preuves Coq est de plus en plus utilisé pour prouver d'importants résultats mathématiques, comme le théorème fondamental de l'algèbre dans la bibliothèque CoRN (Constructive Coq Repository at Nijmegen), le théorème des quatre couleurs, la classification des groupes finis (théorème de Feit-Thompson). Mais Coq est aussi un outil pour formaliser et vérifier d'autres systèmes, par exemple un compilateur pour le langage C, une version légère du langage de programmation Java, ou encore un micro noyau de système d'exploitation. Ces derniers travaux sont particulièrement importants pour les applications industrielles dans des domaines critiques, tels que l'avionique, la monétique et le nucléaire.

III. Programmation et architecture logicielle, systèmes et réseaux

Les individus, les systèmes (au sens large) et les logiciels sont aujourd'hui très fortement

connectés, presque en permanence et presque partout, y compris pendant les déplacements. Cela implique une quantité immense de dispositifs reliés entre eux (serveurs, machines fixes, ordinateurs portables, téléphones mobiles, objets), avec une explosion du nombre de connexions réseau et de données à transférer et stocker. Les défis qui se posent dans un tel contexte sont le passage à l'échelle, la gestion de l'hétérogénéité et la forte dynamique.

A. Nouveaux défis

Dans le domaine des réseaux, des systèmes distribués, du parallélisme et du génie logiciel, les évolutions récentes ont notamment porté sur la mise en œuvre de systèmes adaptatifs et reconfigurables, sur l'introduction de robustesse dans ces systèmes devenus de plus en plus critiques, sur la mise en œuvre de systèmes ouverts afin d'obtenir des solutions plus efficaces, sur l'exploitation du parallélisme et du distribué de masse (avec notamment les multicœurs et le cloud), et l'observation des systèmes déployés et de leurs usages.

Ces différentes problématiques ont soulevé des questions communes, comme la mise en œuvre de modèles, dès la phase de conception des systèmes, qui sont capables de prendre en compte les aspects adaptatifs, distribués, hétérogènes, les usages, et qui passent à l'échelle, ainsi que le déploiement d'expérimentations afin de comprendre et d'évaluer les solutions proposées dans un contexte réaliste.

B. Auto-adaptation

L'adaptation des systèmes au contexte dans lequel ils sont plongés a été un des challenges de ces dernières années. De par la multiplicité, la complexité et l'hétérogénéité des contextes, il est aussi important que cette adaptation se réalise de manière autonome sans besoin d'un paramétrage humain. Parmi les recherches menées dans ce domaine, on peut par exemple noter, dans le système LTE, une allocation dynamique des ressources radio aux mobiles en fonction de l'environnement radio et du trafic qui permet d'améliorer le débit des applications qui en ont besoin et qui se trouvent dans un contexte radio favorable. De plus en plus d'applications s'adaptent maintenant au système sur lequel elles fonctionnent ainsi qu'au réseau qui véhicule les données entre serveurs et clients. Les nouveaux systèmes d'échange de fichiers pair-à-pair, par exemple, prennent en compte la localisation des pairs dans le réseau, mais aussi leurs performances de connexion, leurs comportements, ou encore les similitudes entre contenus.

Pour pouvoir répondre à ces besoins d'adaptation, les solutions proposées sont très souvent logicielles et cherchent à circonscrire les parties matérielles. Les solutions logicielles ont l'avantage d'être flexibles, plus facilement adaptables et améliorables. Une telle approche a été fortement poussée ces dernières années, notamment dans le domaine des réseaux afin de pouvoir plus facilement s'affranchir des spécificités des équipementiers avec, par exemple, les travaux autour de Network Functions Virtualisation, de Software Defined Network ou de Software Defined Radio. De telles propositions rendent également les expérimentations plus facilement envisageables et réalisables.

Avec l'évolution des technologies informatiques, de plus en plus d'informations sont récoltées et de plus en plus d'analyse et de déduction peuvent être tirées de ces données. Ces quatre dernières années ont encore vu s'amplifier cette tendance avec l'application réaliste de techniques de parallélisme étudiées depuis plus longtemps. Ainsi, les réseaux (physiques ou virtuels, voire mixtes), les grappes, les grilles, mais aussi les architectures internes des ordinateurs eux-mêmes, incluant désormais du parallélisme généralisé, ont ouvert de nouvelles perspectives qui étaient jusque-là hors d'atteinte. Les domaines d'application sont extrêmement variés, allant du domaine de la santé au cinéma ou à l'image en général, en passant par l'astronomie, la climatologie, la sociologie, la cryptographie, les statistiques ou encore l'archivage et la documentation.

C. Études et mesures

Les grands systèmes informatiques déployés par l'homme (comme par exemple l'Internet, le cloud, les grands logiciels, ou les systèmes d'exploitation) induisent une complexité qui échappe même à leurs concepteurs. À l'instar des sciences du vivant, des sciences physiques, et des sciences humaines et sociales, l'informatique fournit alors des éclairages précieux en observant ces systèmes complexes par des mesures de leurs comportements, de leurs structures, et des usages qui en sont faits. Ces pratiques ne sont pas fondamentalement nouvelles, puisqu'elles existent depuis les débuts de la discipline, mais cette approche s'est généralisée ces dernières années et joue maintenant un rôle clé dans toutes les facettes de la recherche sur les systèmes, réseaux et logiciels. On analyse ainsi des codes sources, des traces d'exécution, et même des exécutables ; on mesure l'Internet (son infrastructure comme le trafic acheminé) ; on étudie la mobilité (des individus, des véhicules, etc.), les attaques, etc. La France est très présente sur les mesures et l'analyse de la mobilité (dans les hôpitaux, les réseaux véhiculaires, etc.) ainsi que de l'Internet (trafic, infrastructure, usages), avec par exemple les plate-formes SensLab, MOSAR, le LHS ou OneLab.

En parallèle de ces grands défis, la maîtrise de la complexité des systèmes de traitement de l'information et de communication, due notamment à une large échelle et à une dynamique de plus en plus présente (induite par la mobilité des composants, leur défaillance et leur intégration et retrait continus), constitue un enjeu majeur d'autant plus important que ces systèmes sont conduits à supporter des services et applications de plus en plus critiques. Cette complexité rend nécessaire l'utilisation de modèles permettant de comprendre les propriétés fondamentales de ces systèmes et de les concevoir selon une démarche formelle permettant d'en maîtriser les comportements. Un nouveau défi, apparu ces dernières années, pour ces modèles, est qu'ils doivent à la fois spécifier des comportements discrets (à la manière des automates) et des contraintes continues (liées au temps, à la consommation énergétique par exemple, et que l'on résume en contraintes non fonctionnelles). La recherche française (et plus généralement la recherche européenne) est traditionnellement bien positionnée dans ce domaine des méthodes formelles, avec le développement d'outils comme, par exemple, Papyrus.

D. Expérimentations

La validation des solutions proposées est une étape importante dans les domaines des réseaux, des systèmes distribués et des logiciels. Les approches classiques pour une telle validation sont les analyses théoriques, les simulations, l'émulation et/ou les expérimentations. Si l'évaluation théorique et la simulation constituent toujours des sujets de recherche très actifs, avec une implication forte de la France dans le développement de simulateurs comme NS3 et SimGrid, ces dernières années ont vu un développement important autour des expérimentations avec, notamment, la mise à disposition auprès des chercheurs de plates-formes d'expérimentation et d'outils d'expérimentation ouverts. La France est au premier plan sur ces aspects avec par exemple, la plate-forme d'expérimentation FIT dédiée à l'Internet du futur et des objets et la plate-forme Grid'5000 principalement dédiée au calcul haute performance et aux systèmes distribués.

IV. Données et connaissances – intelligence artificielle et interactions

L'intelligence artificielle a connu des évolutions remarquables ces dernières années. Les

laboratoires français ont eu des contributions significatives dans les domaines de l'apprentissage, de la fouille de données et de la théorie de la décision.

A. Apprentissage

Le souci des applications, et en particulier l'exploitation des bases de données, a focalisé l'intérêt des chercheurs sur l'apprentissage de motifs ou de régularités au sein de données. Le but est de découvrir un modèle, celui-ci pouvant prendre la forme d'une théorie ou d'une distribution de probabilités par exemple. Or l'espace des fonctions des modèles est immensément vaste. Dès lors l'inférence d'un modèle particulier à partir de données disponibles pose la double question de l'exploration de cet espace et de l'estimation de la qualité des modèles envisageables par cette exploration. Il faut donc, d'une part, définir un critère inductif permettant de jauger les modèles en fonction des observations et de toute autre connaissance préalable, et d'autre part, trouver un moyen de guider une recherche dans l'espace des modèles. C'est ce double défi qui est au cœur de l'apprentissage et celui qui motive les algorithmes développés.

Un nombre croissant de champs d'applications (génomique, analyse de texte, recherche d'information sur le web, étude de réseaux sociaux...) soulèvent des problèmes qui ne s'inscrivent pas dans le cadre classique de l'apprentissage et nourrissent de nouvelles directions de recherche :

– Données non indépendantes et identiquement distribuées. L'apprentissage ne peut plus se faire en une fois, à partir d'une base fixée d'exemples. Il faut avoir recours à des algorithmes d'apprentissage en ligne dont la complexité devient alors un enjeu majeur.

– Données non vectorielles. Les données sont alors de formats différents ce qui pose des problèmes de comparaison et de définition de la notion de distance.

– Graph mining, c'est-à-dire l'exploitation de données relatives à des graphes. L'objectif peut être de caractériser les nœuds d'un graphe, ou de comparer des graphes. Dans tous les cas, de nouvelles techniques de comparaison de données aux hypothèses doivent être mises au point.

Une autre problématique vient de l'apprentissage en présence de grandes masses de données. Dans cette situation, le temps de calcul devient le principal facteur limitant les performances de l'apprentissage statistique. Cette considération est à contre-courant de la théorie de l'apprentissage statistique traditionnelle qui prend rarement en compte le coût des algorithmes d'apprentissage. Alors que les travaux de Vapnik ne s'y intéressent pas, ceux de Valiant n'autorisent que des algorithmes de coût polynomial. Les travaux fondateurs de Bottou (Microsoft Research, Redmond) et Bousquet (Google, Zurich) démontrent que ce changement d'échelle conduit vers un compromis qualitativement différent, et que par conséquent le meilleur algorithme d'optimisation n'est pas nécessairement le meilleur algorithme d'apprentissage. Par exemple, bien que la descente de gradient stochastique soit un algorithme d'optimisation médiocre, ils montrent, en théorie et aussi en pratique, que sa performance est excellente pour l'apprentissage à grande échelle. Du point de vue logiciel, on assiste à un important développement des systèmes de recommandation, comme ceux qui interviennent dans la gestion du contenu des pages web : le succès fulgurant de Criteo ou le système de recommandation utilisé par Amazon en sont des exemples frappants.

B. Masse de données et de connaissances

Avec l'avènement du Web sémantique et du Linked Open Data, les données et connaissances disponibles dans de nombreux domaines d'application (biologie, médecine, physique, journalisme, culture, loisir) et accessibles via Internet

constituent un énorme gisement de connaissances à découvrir et à valoriser. La nécessité d'intégrer, de croiser, d'interroger à grande échelle des données distribuées et hétérogènes a mis en avant deux grandes thématiques à la croisée des bases de données et de la représentation de connaissances formalisées pour le raisonnement automatique.

La première concerne l'exploitation d'ontologies pour la structuration, l'intégration et l'interrogation de données à grande échelle. La communauté française a ici contribué activement au raisonnement à grande échelle sur des données incomplètes en présence d'ontologies. Cette problématique a forcé les chercheurs à concevoir des logiques de description d'ontologies ayant de bonnes propriétés en complexité des données, et à optimiser les algorithmes de raisonnement et d'interrogation des données en présence d'ontologies par des techniques de réécritures de requêtes. Ces travaux ont fortement influencé les recommandations récentes du W3C pour les langages standards du Web sémantique et ont permis une large adoption des principes du Linked data qui sont en train de révolutionner le Web, où fleurissent de façon décloisonnée données et ontologies interconnectées.

La deuxième se consacre à la fouille de données pour l'extraction automatique de connaissances. Ici aussi, la communauté française a obtenu plusieurs contributions significatives. Des progrès importants ont permis un meilleur passage à l'échelle d'algorithmes avancés en fouille de données pour l'extraction de motifs fréquents ou de règles d'association à partir de données massives de plus en plus complexes (séquences, arbres, graphes), à la fois par l'exploitation du parallélisme mais aussi par des structures de données et des techniques d'exploration efficace issues du rapprochement avec l'analyse formelle de concepts.

C. Théorie de la décision

La théorie mathématique de la décision individuelle modélise le comportement d'un agent face à des situations de choix. Jusqu'à récemment, les travaux portaient sur le développement de modèles axiomatiques dans la lignée des travaux de Savage. Étant donnés un ensemble de choix, une distribution de probabilités sur les états possibles et une fonction d'utilité, l'idée est de définir des axiomes qu'une règle de décision devrait satisfaire. Plus récemment, l'attention s'est tournée vers les questions algorithmiques de la décision. La communauté française a eu des contributions saillantes dans ce domaine et a créé l'International Conference on Algorithmic Decision Theory.

D'autre part, après le succès des systèmes multi-agents tant au niveau théorique qu'appliqué, nous assistons à l'émergence d'une nouvelle problématique, appelée choix social, où un groupe d'agents doit s'accorder de manière collective sur une décision commune selon une procédure centralisée. Les applications variées concernent aussi bien le vote, que le partage équitable de ressources, ou la recherche de consensus (agrégation de jugement). Dans les deux premiers cas, il s'agit d'agréger des préférences, alors que dans le troisième cas il s'agit d'agréger des croyances. À la croisée entre théorie du choix social et informatique se développe un nouveau domaine de recherche, le choix social computationnel, dont un des objectifs est d'importer des concepts et procédures de la théorie du choix social pour résoudre des problèmes issus de l'informatique (par exemple les procédures d'agrégation pour le classement de pages web et la recherche d'information, ou encore l'utilisation de procédures de vote pour la classification et la reconnaissance des formes), un autre but étant d'utiliser des notions et méthodes venant de l'informatique (langages de représentation, complexité, algorithmique, protocole d'interactions...) pour résoudre des problèmes de décision de groupe complexes.

V. Aide à la décision et recherche opérationnelle

La recherche opérationnelle et l'aide à la décision (RO-AD) forment une discipline par nature interdisciplinaire, couvrant plusieurs champs scientifiques comme les mathématiques, l'informatique, la gestion et l'économie. Cette discipline a pour objet l'étude générale des problèmes de décision (modéliser, analyser, prévoir, optimiser, simuler, évaluer, etc.) issus de contextes applicatifs. Elle a pour vocation non seulement d'obtenir des résultats fondamentaux mais également de fournir des réponses concrètes aux problèmes complexes posés en pratique, notamment sous la forme de logiciels. Au-delà des systèmes de production, de la logistique industrielle, ou de la conception et l'exploitation des grands réseaux physiques (transport, énergie, télécommunications), depuis la fin des années 90, le champ d'application de la RO-AD s'est étendu à de nouveaux secteurs tels la logistique hospitalière, la génétique, la conception de circuits micro-électroniques, l'ordonnancement d'architectures de calcul massivement parallèles, la robotique, et la finance.

En France, la RO-AD est historiquement centrée sur les problèmes d'optimisation mathématique et rejoint ainsi naturellement les mathématiques discrètes et appliquées, ainsi que l'informatique fondamentale.

A. Heuristiques et modèles

Dans ce contexte, nous observons deux tendances fortes. La première est relative à l'approximation, au caractère heuristique des approches algorithmiques. Les modèles s'enrichissent et leur échelle augmente à mesure que les succès de la discipline se déploient dans le monde économique. La masse exponentiellement croissante des données à disposition tend à renforcer cet effet. Simultanément les besoins en réactivité se sont accrus, limitant souvent les temps de résolution à quelques minutes, parfois quelques secondes. Ainsi, nous assistons depuis une dizaine d'années à l'avènement des heuristiques, pour lesquelles on s'emploie à caractériser les temps d'exécution, au regard de la qualité des solutions produites. Cette tendance s'observe autant en optimisation combinatoire avec les algorithmes dits de recherche locale et les algorithmes évolutionnaires, qu'en optimisation continue avec les algorithmes à base de gradient approché voire stochastique. Ces approches algorithmiques heuristiques itératives permettent aujourd'hui de traiter de façon satisfaisante en pratique des problèmes de très grande taille (jusqu'à des millions de variables de décision dans certains cas). Néanmoins, leur caractère heuristique, hybride (c'est-à-dire mélangeant plusieurs paradigmes algorithmiques), itératif, souvent randomisé et parallélisé (voire distribué), les rend extrêmement complexes à analyser que ce soit dans le pire des cas ou plus encore pour des instances réalistes. Les heuristiques avec garantie de performance permettent de pallier ce défaut mais n'offrent généralement pas un bon comportement moyen. Un défi dans ce domaine reste donc de parvenir à de bonnes heuristiques en pratique qui ont aussi des garanties de performance.

La seconde tendance concerne les modèles de décision et d'optimisation eux-mêmes, qui initialement déterministes et simplistes sont devenus plus riches. Des modèles d'optimisation stochastique traitant les aléas dans les données du problème ont été développés, ainsi que des modèles d'optimisation robuste visant à gérer les risques dans la prise de décision. Il est également davantage fait appel à la notion de ré-optimisation, qui permet d'appréhender le caractère dynamique des contextes d'optimisation sous un angle plus opérationnel. Bien que peu mis en œuvre jusqu'à présent, les modèles d'optimisation multicritère connaissent un regain d'intérêt. Ils s'inscrivent dans un courant plus large, visant à étudier et intégrer les modèles produits par les sciences humaines et sociales, permettant de prendre plus finement en compte les facteurs humains et de situation dans la prise de décision. Ce

courant rejoint ainsi les préoccupations d'une partie de la recherche effectuée en intelligence artificielle.

B. CSP et SAT

Les modèles de satisfaction se rencontrent principalement dans le paradigme de la programmation par contraintes (CSP) et la satisfaisabilité de clauses logiques (SAT). Les travaux menés dans ce contexte ont été particulièrement intenses ces dernières années à la fois sur les plans théorique et pratique, et ont été associés au développement de solveurs de plus en plus performants. Citons en premier des résultats théoriques importants faisant le pont entre toute une famille de résultats d'inapproximabilité et la conjecture des jeux uniques posée par Khot (une variante plus puissante du théorème PCP), de nouveaux résultats de dichotomie sur de nouvelles classes de CSP (ces CSP sont alors soit dans P soit NP-complets), ou encore sur les phénomènes de seuils de SAT.

Les approches algorithmiques ont vu l'avènement de méthodes d'optimisation convexe pour des problèmes de très grande taille présentant d'excellentes garanties de convergence. En optimisation robuste, où là aussi on s'intéresse à des problèmes de très grande taille, des travaux récents permettent de garantir certains types de complexité algorithmique en fonction du modèle d'incertitude utilisé. Ces avancées ainsi que le raffinement des modèles se sont concrétisés par un déploiement massif de logiciels d'optimisation dans les entreprises de tous les secteurs et de toutes les tailles, avec même une accélération ces dernières années. Ils sont entre autres devenus un outil indispensable du praticien en RO-AD. Tirée par des besoins économiques forts, la tendance est à l'hybridation voire l'unification des différentes techniques d'optimisation au sein de solveurs de programmation mathématique toujours plus génériques, aussi appelés solveurs d'optimisation globale, qui apparaissent aujourd'hui comme le nouveau graal de la discipline. Encore à l'état de prototypes, un challenge scientifique et technologique est de parvenir à rendre ces solveurs opérationnels.

Ces avancées ont aussi eu un impact fort sur plusieurs domaines en intelligence artificielle comme la planification d'actions où les nouveaux planificateurs utilisent des solveurs SAT. Les langages de programmation logique ont aussi bénéficié de ces travaux, notamment ASP (« Answer Set Programming »). Aujourd'hui, par la disponibilité de plusieurs solveurs performants, l'ASP apparaît comme une implantation effective du raisonnement non monotone comme il avait été théorisé par Reiter en 1980 en logique des défauts. Mais bien au-delà du raisonnement non monotone, l'ASP est également un formalisme adapté à de nombreux domaines de la représentation des connaissances en intelligence artificielle (raisonnement de sens commun, web sémantiques…) et à la résolution de problèmes combinatoires (planification, problèmes de théorie des graphes, configuration, bioinformatique...). Ce dernier aspect, certainement le plus prometteur pour l'ASP, est basé sur l'idée de coder chaque problème à l'aide d'un programme logique dont les modèles correspondent aux solutions du problème.

C. Rayonnement et transfert

La RO-AD française jouit d'une bonne visibilité internationale dans le domaine de l'optimisation, discrète ou continue. Contribuant peu aux avancées en programmation linéaire en variables mixtes, elle se distingue en optimisation combinatoire, optimisation convexe et en optimisation globale. Elle est également à la pointe dans le domaine des techniques SAT et à base de contraintes. Les activités de transfert sont bien développées. De nombreuses grandes entreprises et institutions françaises disposent ainsi de compétences voire d'un département de RO-AD : Airbus, Air France, Air Liquide, Bouygues, Dassault Aviation, EDF, Google France, GDF SUEZ, Miche-

lin, Orange, Renault, SNCF, Thales, Veolia, etc. La France compte également un réseau de plusieurs dizaines de PME et TPE fournissant des services spécialisés dans le domaine. Enfin citons le cas de la société française ILOG, leader mondial dans l'édition de logiciels en RO-AD, qui a récemment été acquise par IBM.

VI. Nouvelles interactions avec les autres sciences

Nous présentons dans ce qui suit uniquement les nouvelles interactions avec les sciences sociales, biologiques et physiques. Celles avec les mathématiques, d'une nature différente et plus diffuse, ont été pour la plupart présentées dans les sections précédentes. L'apport de l'informatique aux autres sciences a débuté avec le traitement informatique des données produites par les autres sciences. Il serait toutefois réducteur de se limiter à ce type d'interaction, puisque l'informatique, en tant que science à part entière, permet de mieux comprendre certaines grandes questions scientifiques grâce aux notions et concepts qui lui sont propres. Inversement, l'informatique s'inspire elle-même des autres sciences pour proposer et étudier de nouveaux modèles de calcul.

A. Du traitement de données scientifiques vers les sciences computationnelles

Le traitement informatique de la masse des données scientifiques récoltées dans tous les champs de recherche est une source historique d'interactions entre disciplines, qui a rencontré de grand succès, comme en témoignent par exemple les appels Mastodons lancés par la Mission pour l'Interdisciplinarité du CNRS. Au-delà de ce phénomène, ces dernières années ont vu un changement qualitatif fondamental avec la naissance de nouvelles problématiques. En sciences humaines et sociales (SHS), par exemple, les grands corpus de données (banques d'images, films, enquêtes) sont désormais systématiquement numérisés, et l'acquisition de données totalement nouvelles ouvre de nouveaux champs sur leur usage (ainsi les smartphones constituent un formidable réseau de capteurs). En sciences du vivant, l'accessibilité des technologies de séquençage pose de nouveaux défis, à la fois en termes d'exploitation de la masse de données collectées mais aussi en termes de fiabilité des outils. La biologie cellulaire a subi une mutation similaire avec l'acquisition haut-débit d'images de fluorescence qui produit un flux de données extrêmement important permettant un saut qualitatif dans les questionnements abordés puisque des résolutions spatio-temporelles extrêmement fines sont désormais accessibles sur cellule, voire molécule unique. Les nouvelles problématiques de traitement de données qui émergent ici rappellent d'ailleurs celles connues depuis plus longtemps en astrophysique ou en physique des particules.

Depuis peu se développe un nouveau type d'interactions, plus riche, qui utilise l'approche informatique pour aborder des questions scientifiques plus larges que le traitement de données. Dans cette perspective, l'informatique n'est plus vue comme un outil de calcul mais comme une nouvelle approche à part entière, munie d'une multitude de modèles généralement discrets (graphes, réseaux de Petri, automates, algorithmes...) et des notions de complexité qui s'y rapportent. Ce courant scientifique se développe très rapidement et permet par exemple d'introduire une notion de complexité (relative à un algorithme, un protocole, un système distribué...) dans d'autres disciplines.

Les travaux fondateurs de Valiant permettent ainsi de mesurer quantitativement les performances des recombinaisons génétiques dans certains modèles d'évolution. Toujours en biologie, une autre thématique importante se développe autour de démarches mixant approches expérimentales et modélisation, souvent

regroupées sous l'intitulé de biologie des systèmes mais incluant aussi des approches de neurosciences computationnelles. Une étape majeure a été franchie en 2012 avec la publication du premier modèle computationnel d'une cellule bactérienne complète *(Mycoplasma genitalium)*. Si la France a obtenu dans ces domaines quelques grandes réussites, force est de constater qu'elle reste relativement en marge de ce courant pourtant majeur à l'international.

En physique, en chimie et en biologie structurale, la simulation est aujourd'hui une composante indispensable de la progression des connaissances. Un exemple particulièrement marquant est celui de l'attribution du prix Nobel de chimie en 2013 à trois pionniers de la chemo-informatique ou la résolution, par des approches de combinatoire, de problèmes ouverts en physique statistique comme le problème PASEP. Encore en physique, le développement de l'informatique quantique, outre ses aspects algorithmiques et cryptographiques, entraîne aussi de nouveaux éclairages et questionnement sur la notion d'intrication quantique, comme la loi d'aire mêlant intrication et théorie de l'information.

En économie enfin, la théorie des jeux a fait émerger des notions telles que les mécanismes de marchés (mechanism design), ou les systèmes de recommandations qui ont radicalement changé la compréhension et l'interprétation de certains phénomènes économiques.

B. Nouveaux modèles de calculs inspirés des autres sciences

De très nombreux systèmes réels peuvent être vus comme des systèmes de calcul. C'est le cas des systèmes quantiques, du cerveau et des systèmes nerveux mais aussi des systèmes sociaux (par exemple, certains groupes d'insectes). Utiliser ces systèmes pour proposer de nouveaux modèles de calcul constitue une discipline à part entière de l'informatique.

Historiquement, ce courant est né à l'interface de l'informatique et des sciences du vivant avec le développement de méta-heuristiques bio-inspirées (réseaux de neurones, algorithmes évolutionnaires...). Un de ces grands courants qui connaît une très forte activité sur ces dernières années est le deep-learning, qui profite directement du développement de très grands corpus de données peu ou pas structurées, et repose sur des architectures typiquement inspirées des neurosciences.

L'informatique quantique est sans doute le modèle de calcul qui révolutionne le plus notre compréhension du traitement de l'information. Un effort très important continue d'être réalisé par une communauté regroupant informaticiens, physiciens, mathématiciens et ingénieurs afin de s'attaquer aux problématiques majeures du domaine : algorithmique et complexité quantique, cryptographie quantique, codes correcteurs d'erreurs et réalisation à grande échelle d'ordinateurs quantiques. Une préoccupation nouvelle des informaticiens concerne l'implantation. Ont ainsi vu le jour des modèles de calcul plus réalistes mais restant supérieurs aux ordinateurs traditionnels (boson sampling), ainsi que des réalisations de nouveaux protocoles cryptographiques, de chiffrement à plusieurs centaines de kilomètres de distance, et de systèmes quantiques clé en main. Dans toutes ces directions, la France a apporté plusieurs contributions phares. La création en 2014 de la fédération de recherche Paris Centre for Quantum Computing (PCQC) va encore améliorer sa visibilité et la confirmer en tant que point névralgique de la recherche en informatique quantique à l'international.

Motivé par le développement rapide des approches d'ingénierie du vivant, la notion de calcul moléculaire progresse en relation avec la biologie synthétique. Les progrès importants de ces dernières années permettent d'envisager l'implantation d'éléments mémoire et de portes logiques au sein même d'un chromosome bactérien. Si l'ordinateur bactérien est encore loin, ces avancées pourraient rapidement permettre le développement de capteurs bactériens, éventuellement organisés en réseaux.

Conclusion

La communauté française bénéficie d'une grande visibilité internationale et a obtenu récemment toute une série de résultats remarquables, notamment sur des sujets fondamentaux aux répercussions pratiques parfois essentielles. Cette caractéristique de la recherche en informatique en France est une force qu'il convient de soutenir. Il s'agit d'un enjeu stratégique tant les répercussions à long terme sont difficilement prévisibles, mais néanmoins souvent importantes, et assez directes notamment dans la plupart des thématiques de l'informatique.

Certaines interactions avec les autres disciplines gagneraient à être renforcées, par exemple celles avec les SHS et en particulier avec l'économie, afin de ne pas rater certaines grandes évolutions de l'informatique. Il s'agit ici de développer de fructueux échanges où l'informatique, en tant que science à part entière, peut apporter ses méthodologies et ses paradigmes et en retour s'inspirer de ceux issus d'autres disciplines.

Le renforcement de certaines thématiques émergentes, dont les avancées sont conditionnées à des collaborations fortes entre chercheurs issus de communautés distinctes pourrait être réalisé grâce à la mise en place de structures de recherche, éventuellement temporaires. Le Laboratory of Information, Networking and Communication Sciences (LINCS) réunit ainsi à Paris chercheurs du monde académique et du milieu industriel issus du domaine des réseaux. Un autre exemple de structure, entièrement académique cette fois, tout à fait remarquable et dont la communauté et ses tutelles pourraient également s'inspirer est celui du Center for Massive Data Algorithmics (Madalgo) à Aarhus (Danemark) co-financé par plusieurs institutions au Danemark et en Allemagne, ainsi que par le MIT.

Enfin, l'effort de diffusion et valorisation des résultats, et plus particulièrement des logiciels et des outils développés doit être poursuivi et amplifié. Cela nécessite un soutien des tutelles notamment par le recrutement d'ingénieurs et techniciens en nombre suffisant.

Comité national de la recherche scientifique. « Section 06- Sciences de l'information : fondements de l'informatique, calculs, algorithmes, représentations, exploitations ». *Rapport de conjoncture 2014*, [édition PDF en ligne]. ISBN : 978-2-271-08746-1. Disponible sur : http://rapports-du-comite-national.cnrs.fr/