

AXES DE RECHERCHE

-

DOMAINE D'INTERET MAJEUR LOGICIELS ET SYSTEMES COMPLEXES

Axes de recherche

L'activité du DIM LSC concerne la méthodologie de la conception et le développement de systèmes à forte composante logicielle. L'informatique dans toutes ses dimensions, mais aussi l'approche systèmes de l'ingénierie, les notions d'automatique et de traitement du signal et des images, les réseaux et les télécommunications y jouent un rôle prépondérant. Parmi les systèmes concernés, mentionnons les systèmes intégrés sur une puce (*systems on chips*), les systèmes de calcul à haute performance, les infrastructures des grands logiciels, les systèmes de télécommunication, les systèmes embarqués et les robots.

Ces systèmes complexes présentent les caractéristiques suivantes :

- ils doivent interagir avec le monde extérieur (capteurs et actionneurs intelligents) et avec leurs utilisateurs (activation vocale, interfaces haptiques, visualisation, environnements interactifs),
- leur développement passe par la modélisation (et éventuellement la simulation) afin d'aider à leur conception, à leur optimisation, à la prédiction de leurs performances et en particulier de leur robustesse,
- ils requièrent une connaissance approfondie des fondations théoriques et des aspects méthodologiques du développement de logiciel et de ses liens avec le matériel, pour tenir compte de critères comme la correction, l'efficacité, la robustesse et la fiabilité,
- ils posent enfin souvent des problèmes délicats de gestion de masses de données gigantesques et disséminées, ainsi que d'utilisation d'une puissance de calcul répartie sur des réseaux d'ordinateurs.

Les études soutenues doivent présenter un certain caractère de généralité dans le domaine du logiciel ou du système. Le fait, par exemple, qu'un projet comporte l'écriture d'un code informatique dédié ne suffit pas à garantir sa recevabilité. Une étude système de nature très théorique ne se traduisant pas par l'écriture d'un code pourra par contre s'avérer recevable.

Le DIM LSC encourage les équipes de la région parisienne qui travaillent sur ces thématiques à tirer profit de leur diversité et de leur complémentarité pour consolider leurs compétences et les exploiter pour résoudre ensemble des problèmes plus efficacement qu'elles ne le feraient de façon isolée et pour préparer le futur.

Les paragraphes qui suivent ne sauraient bien sûr constituer une liste exhaustive de domaines dignes d'attention, et le comité du DIM LSC restera ouvert à d'autres propositions.

1- Logiciel

Le DIM LSC s'intéresse aux problèmes posés par le développement, la validation et la vérification des logiciels. Le but est de rendre ces activités plus sûres, pour obtenir des logiciels de meilleure qualité, avec un processus de production plus efficace. La recherche dans ce domaine va des fondations théoriques (sémantique, systèmes à types) jusqu'au développement d'outils pratiques. Dans tous les cas, le défi principal est de permettre le traitement de problèmes de plus grande taille.

Le développement, la vérification et la validation de logiciels reposent sur des modèles et des spécifications. Les systèmes visés sont en particulier les systèmes temps-réel embarqués et les systèmes distribués. Les approches de modélisation vont du pragmatique au formel.

La production de systèmes sûrs à base de logiciels est un défi auquel des efforts considérables sont consacrés. Des recherches sur la validation et la vérification doivent être conduites tant au niveau des modèles qu'à celui des programmes. Il existe déjà un noyau de travaux sur des outils génériques dans les domaines de la preuve automatique de théorèmes, de la résolution de contraintes, de la cryptographie, et de la combinatoire, qui requièrent une recherche fondamentale en informatique et en mathématique. Des projets plus appliqués concernent l'analyse statique et la preuve de programmes, et la vérification de modèles probabilistes. La validation des approches proposées passe par le traitement de problèmes significatifs et de projets industriels. Le passage à l'échelle est alors un aspect très important à prendre en compte, et les méthodes de vérification approximative et les approches statistiques sont prometteuses à cet égard.

Les architectures tolérantes aux fautes et le diagnostic en ligne sont également à considérer.

Frontières du logiciel avec le matériel

Les sujets d'intérêt vont de l'optimisation de compilateurs pour des processeurs embarqués au développement d'outils logiciels pour la simulation rapide de processeurs en passant par le développement et la certification de systèmes d'exploitation temps réel sûrs.

Ingénierie des connaissances, concepts et outils de fouille de données

Les bases et entrepôts de données, la modélisation de connaissances, l'utilisation d'ontologies dans les systèmes de recherche d'information relèvent de cette thématique, avec des applications aux services web. Un des défis consiste à fournir un accès facile aux sources pertinentes. Quand celles-ci sont nombreuses et hétérogènes, ceci suppose le développement d'outils automatiques d'annotation et de recherche. La fouille de données utilise des techniques d'analyse statistique et d'apprentissage.

2- Modélisation et simulation

La modélisation (qu'elle repose sur une connaissance précise des phénomènes physiques impliqués ou pas) et la simulation numérique jouent un rôle stratégique pour la conception et l'évaluation de systèmes. La simulation remplace même de plus en plus les tests sur les systèmes physiques. Elle joue également un rôle clé dans l'optimisation de l'expérimentation pour la qualification des systèmes.

La simulation numérique s'applique pratiquement à toutes les industries où la compétitivité est fortement liée à la durée des cycles de conception et de développement, à la gestion de risque ou à la réduction de l'impact écologique. Elle peut permettre de réduire de façon drastique les coûts de conception et de maintenance, le temps de mise sur le marché et l'impact du changement de la réglementation. Les programmes de simulation numérique deviennent de plus en plus complexes, pour prendre en compte des phénomènes physiques multiples, à des échelles variées. Ils doivent exploiter la puissance de calcul toujours croissante, et mobiliser si nécessaire les systèmes de calcul à haute performance ou le calcul sur grille. Ils doivent souvent manipuler de très grandes masses de données.

Le DIM LSC est notamment intéressé par les sujets suivants :

Méthodes numériques : Réduction et optimisation de modèles, adaptation de maillages, décomposition de domaines, quantification des erreurs numériques, méthodes à éléments finis, généralisation de méthodes initialement développées pour des applications spécifiques...

Modélisation numérique multiphysique : Les physiques concernées peuvent être, par exemple, la neutronique, la thermodynamique, l'hydraulique, la mécanique, la chimie, la métallurgie. Un des défis est d'interfacer des composants logiciels traitant des physiques différentes tout en ne détériorant pas leurs performances et sans créer des discontinuités artificielles.

Modélisation numérique multiéchelle : Les échelles concernées peuvent être *temporelles* (de la nanoseconde au siècle en ce qui concerne le stockage des déchets nucléaires, par exemple) ou *spatiales* (de la structure macroscopique aux constituants atomiques, par exemple).

Adaptation du logiciel au matériel : Il peut s'agir d'équilibrer la charge de calcul entre plusieurs unités de calcul ou de la conception d'algorithmes spécifiques à mettre en œuvre sur des grilles ou des systèmes de calcul à haute performance.

Supervision de codes : Les préoccupations concernent notamment les pré- et post-traitements adaptés aux maillages de très grande taille (avec des millions de degrés de liberté), les stratégies de couplage de codes, l'optimisation, la construction de modèles simplifiés par apprentissage, le traitement des incertitudes (modélisation et propagation) et les méthodes probabilistes.

3- Calcul

L'exploitation du parallélisme au niveau des instructions et l'accroissement de la fréquence d'horloge ont atteint leurs limites en ce qui concerne l'amélioration des performances de calcul. Depuis maintenant plusieurs années, la quête des performances ultimes sur un processeur unique est terminée, et on assiste à un mouvement massif en direction des architectures multi-cœur. L'intégration de dizaines, voire de centaines de cœurs devient envisageable dans un futur proche. Ce nouveau paradigme a un impact profond sur tous les aspects techniques, de l'architecture matérielle aux langages de programmation.

Les défis principaux portent sur la gestion de la complexité qui va avec l'utilisation d'un grand nombre de ressources. Il faut aussi proposer des modèles d'exécution innovants et des solutions pour dépasser les goulots d'étranglement liés aux accès mémoire et aux interconnexions. La gestion de l'énergie et la fiabilité sont également des points clé, comme l'est le développement de langages présentant un haut niveau d'abstraction pour exploiter pleinement le parallélisme sans obliger le programmeur à se plonger dans les aspects architecturaux.

Des domaines de recherche clé sont donc

- les modèles d'exécution et la gestion de la complexité,
- les processeurs de calcul intensif à architecture multi-cœur,
- les réseaux d'interconnexion sur puce et la gestion de la mémoire,
- la compilation prenant en compte les aspects énergétiques et la gestion de l'énergie lors de l'exécution (pour les systèmes embarqués),
- la fiabilité, la sûreté de fonctionnement et la tolérance aux fautes,
- les modèles de programmation et les langages,
- les environnements avancés de développement d'applications dans un contexte multi-cœur.

4- Collecte d'information (instrumentation)

L'approche traditionnelle en instrumentation part d'un capteur (qui produit par transduction un signal lié à la quantité physique à mesurer) pour aboutir à une estimée numérique de la quantité d'intérêt par traitement de ce seul signal.

Une multitude d'autres schémas est maintenant envisageable, grâce à la puissance de calcul disponible et à l'usage massif de la modélisation, de la simulation, des connaissances a priori, des bases de données et des réseaux.

Le concept clé d'*information* et la théorie de l'information doivent être au cœur des systèmes d'instrumentation, qui sont à l'interface entre le monde physique (où les sciences sont la physique, la chimie, la biologie, etc. et où la technologie est celle des capteurs) et le monde numérique (où les sciences sont les statistiques, l'analyse numérique, etc., et où la technologie est principalement logicielle).

Sciences et technologies physiques et numériques doivent s'interconnecter de plus en plus, pour aboutir au concept multidisciplinaire de *système d'instrumentation*, pour lequel le croisement des savoirs et synergie devrait être sources de nouvelles fonctionnalités et de performances accrues.

Parmi les sujets de recherche dignes d'intérêt, mentionnons

- toute approche à base de traitement de l'information qui transforme des données produites par des capteurs (et des modèles et des bases de données) en des estimées des quantités d'intérêt,
- toute approche où la conception du capteur est pilotée par des préoccupations de traitement de l'information,
- les problèmes inverses, couplant des capteurs physiques, des modèles physiques et des simulations,
- les capteurs logiciels, pour estimer des quantités qui ne peuvent pas être mesurées directement à partir des données et des connaissances a priori disponibles,
- la planification d'expériences par des approches statistiques visant à minimiser l'incertitude sur les quantités d'intérêt,
- la théorie de la commande appliquée à l'estimation non linéaire de l'estimation d'état et de paramètres,
- le couplage de bases de données et de connaissances a priori avec des mesures physiques,
- les réseaux de capteurs.

5- Prise de décisions et commande

Il s'agit ici de prise de décisions à partir d'un traitement automatisé des informations disponibles. La *classification* (par exemple pour du diagnostic médical ou la détection et l'isolation de défauts) et la *commande* (c'est à dire le calcul de signaux d'entrée à appliquer à un système pour le forcer à adopter un comportement désiré) sont des exemples de telles décisions. Ces décisions s'appuient en général sur des modèles mathématiques qui doivent être construits soit à partir des lois de la physique, de la chimie, de la biologie, etc., soit à partir de la seule observation des comportements, soit encore par combinaison des deux approches. La modélisation et l'estimation de paramètres (souvent appelée *identification*) jouent donc un rôle important. Il est aussi important de pouvoir inférer, en utilisant les données disponibles et le modèle, les valeurs prises par des variables qui ne sont pas directement accessibles à la mesure mais jouent un rôle crucial dans le processus de décision. Suivant la communauté concernée, on parle d'*observation* ou d'*estimation d'état* ou de *capteur logiciel*. A cause de la nature imparfaite du modèle et des perturbations inévitables, la commande est en général actualisée pour tenir compte du comportement constaté du système, tel qu'il est mesuré par des capteurs. Un point critique est alors d'assurer la *stabilité* du système commandé.

Les outils à mobiliser pour la conception de systèmes de prise de décision et de commande viennent de l'automatique, du traitement du signal et des images, des statistiques et de l'intelligence artificielle. La collaboration entre spécialistes de cultures différentes attaquant les mêmes problèmes sous des angles différents devrait permettre des recherches fructueuses.

En estimation d'état non linéaire, par exemple, automaticiens, traiteurs de signaux et statisticiens peuvent coopérer pour l'étude de systèmes non linéaires dynamiques incertains afin de produire des estimateurs qui prennent explicitement en compte l'incertitude et l'usage prévu pour l'estimée produite.

Dans le domaine de la modélisation comportementale, des tentatives de créer des ponts entre intelligence artificielle, automatique et statistiques devraient être encouragées. En commande, on suppose souvent que le modèle et l'état du système à commander sont disponibles. Les interactions entre la modélisation, l'estimation d'état et la commande dans un contexte de connaissance très partielle du système et de perturbations inévitables méritent plus de recherches.

Des équipes qui ne se considèreraient pas comme travaillant sur la prise de décision et la commande traitent pourtant de systèmes dynamiques incertains, qu'il peut falloir modéliser et sur lesquels il convient de prendre des décisions. D'autres équipes qui développent des méthodologies de prise de décision et de commande n'ont pas les équipements, les compétences et les ressources humaines requises pour traiter de vraies applications de leurs concepts par exemple en biologie, sur les réseaux de télécommunication, sur le calcul sur grille, en physique ou en robotique. Les deux types d'équipes devraient bénéficier de projets d'applications communs.

Les dispositifs de commande peuvent être embarqués sur des systèmes, ou distribués sur un réseau, qui peut lui-même être utilisé pour commander des objets distribués. Les interactions entre l'automatique, les télécommunications et l'informatique devraient jouer là un rôle important.

6- Interaction, visualisation et réalité virtuelle

L'interaction entre des êtres humains et des machines est l'un des domaines de l'informatique qui ont beaucoup évolué ces dernières années. Les progrès sont dus principalement à l'augmentation de puissance des ordinateurs, aux capteurs à base de nouvelles technologies et aux logiciels interactifs. Ils résultent aussi de nouvelles façons de considérer le rôle des ordinateurs dans la vie de tous les jours.

Les avancées sur les dispositifs de visualisation et sur le calcul intensif offrent des perspectives nouvelles en termes de modélisation des systèmes physiques et de visualisation de grandes masses de données. Cette visualisation passe par de nouvelles façons d'interagir avec une machine dans des systèmes de réalité virtuelle.

Ces recherches impliquent l'étude de la relation étroite qui s'établit entre l'homme et la machine durant l'interaction, ce qui suppose des liens avec la psychologie et la psychophysiologie dans un cadre interdisciplinaire. Il faut en effet tenter de comprendre comment les hommes perçoivent leur environnement technique et interagissent avec lui, et de valider les solutions technologiques qui servent de médium pour les interactions avec les utilisateurs.

Parmi les sujets d'intérêt, on peut mentionner

- La géométrie algorithmique,
- les méthodes de conception de systèmes interactifs,
- le développement de nouvelles interfaces et de nouveaux paradigmes d'interaction,
- l'adaptation au handicap,
- l'interaction située avec les objets technologiques dans un contexte d'intelligence ambiante,
- les systèmes de recherche d'information et de dialogue,
- l'ingénierie des connaissances et la recherche d'information multimédia et multilingue,
- l'analyse (et la synthèse) du langage écrit ou parlé, des images et des vidéos,
- l'analyse de documents non structurés,
- les systèmes de questions et réponses intégrant des éléments extra linguistiques,
- les interfaces multimodales et multi-échelles pour la visualisation et l'exploration de grandes masses de données,
- l'interaction haptique, et plus généralement l'interaction avec des systèmes possédant un grand nombre de degrés de liberté,
- la conception assistée par ordinateur utilisant des systèmes virtuels,
- la conception et l'usage d'êtres humains virtuels (avatars),
- la collaboration à distance dans des environnements immersifs.

7- Robotique

La robotique fait appel à des compétences très diverses pour traiter des problèmes d'ingénierie très variés (conception mécanique, perception, planification des actions, exécution des tâches, interfaces multimodales, etc.). Le problème de la sécurité est crucial, puisque hommes et robots doivent pouvoir cohabiter. La preuve que les logiciels feront bien ce pour quoi ils sont prévus en est un aspect important.

Un robot doit pouvoir adapter de façon autonome son comportement à son environnement en tenant compte des tâches qui lui ont été confiées. Dans le contexte du DIM LMSC, ceci suppose de développer des recherches dans trois directions. La première est la *perception*. Il s'agit pour le robot d'utiliser les capteurs disponibles pour se localiser (éventuellement pour explorer son environnement) et pour identifier les objets qui l'entourent. Les outils mobilisés incluent le traitement du signal et des images. Le robot peut également agir pour faciliter l'opération (perception active et mécanismes attentionnels). La seconde direction concerne le *contrôle du comportement* (commande bas niveau des actionneurs, planification de trajectoires, commande à base de vision, interfaces homme-machine). Les outils mobilisés incluent l'automatique et l'apprentissage automatique. La troisième direction est la conception des robots (architecture physique, nouveaux types d'actionneurs et de capteurs, modélisation et identification).

Les applications concernées incluent les drones, les robots mobiles, les voitures, les télémanipulateurs et les interfaces haptiques, les robots humanoïdes, les robots d'assistance aux personnes handicapées, l'assistance au geste chirurgical, etc.

8- Bioinformatique

La génomique, la métabolomique, la modélisation de la structure moléculaire, l'étude des réseaux de régulation, la compression d'images médicales, l'analyse des effets de l'exposition aux radiations électromagnétiques sont une illustration de la variété des domaines qui fournissent des défis de choix à qui s'intéresse aux systèmes complexes. Il faut mobiliser notamment l'informatique (et en particulier les méthodes permettant le traitement de données massives hétérogènes), les mathématiques, le traitement du signal et des images, et l'automatique. La plupart des grands thèmes du DIM LSC sont susceptibles de contribuer, mais ceci ne peut se faire sans un contact fort avec des laboratoires de biologie dans un thème interdisciplinaire par essence.

9- Systèmes hybrides

Il s'agit ici de concevoir et de valider des systèmes qui combinent des parties à temps continu (associées par exemple avec le monde physique) à des parties pilotées par des événements (associées par exemple à des calculateurs embarqués). Le nombre de ces systèmes explose à mesure que de plus en plus de calculateurs sont connectés à des systèmes physiques qu'ils doivent contribuer à rendre plus adaptables, plus sûrs et plus efficaces.

La conception est nécessairement pluridisciplinaire. Elle s'appuie sur une grande variété de modèles mathématiques (équations différentielles, à événements discrets, hybrides), et l'automatique, l'informatique, le traitement du signal et les statistiques y jouent un grand rôle. Le défi principal est de développer et de disséminer des techniques pour améliorer performances et robustesse tout en réduisant le temps et le coût de la conception, du développement, de la validation, de la production et de la maintenance. Comme de nombreux systèmes sont critiques d'un point de vue sécurité (systèmes de transport ou d'assistance médicale, par exemple), la fiabilité est une préoccupation majeure.

10- Réseaux de communication du futur

La recherche vise notamment à augmenter les capacités des réseaux pour permettre la prise en charge de nouveaux trafics et de nouveaux services, pour améliorer la qualité de service en termes de performances et de fiabilité, et pour faciliter la mobilité. Le nombre des technologies d'accès croît très vite (2G, 3G, 3G LTE, 4G, Wi-Fi hotspots, Wi-Max, xDSL, FTTx, câble ou satellite). Un point clé est la généralisation des réseaux qui s'auto-organisent de façon autonome pour créer de nouveaux services centrés sur leurs utilisateurs ou pour réduire le coût opérationnel des services existants.

Les réseaux *ad hoc* sont des exemples typiques de cette situation. Ces réseaux doivent être améliorés pour prendre en compte des notions comme la qualité de service (bande passante, service temps-réel...), la diffusion multiple, la sécurité, etc. Il faudra également travailler sur l'internet du futur (connectivité, autoconfiguration, réseaux tolérants aux retards, etc.).

Un des points clé pour la conception optimale et la gestion des réseaux et pour l'analyse du trafic demeure l'*évaluation de performance*. De nouvelles méthodes (analytiques, algébriques, numériques, approximatives, par bornes...) sont en cours de développement. Les réseaux de grande taille représentent un défi particulier.