

Rapport de Prospective (Horizon 2030)
du
Groupement de Recherche en
System On Chip, Systèmes embarqués et Objets Connectés
(SoC²)

GdR - 2995

<http://www.gdr-soc.cnrs.fr/>

Directeur : Ian O'CONNOR
PR1 : École Centrale de Lyon – INL
ian.oconnor@ec-lyon.fr
Tél : 04 72 18 60 54 ; Fax : 04 78 43 35 93

Directrice adjointe : Cristell MANEUX
PR1 : Université de Bordeaux - IMS
cristell.maneux@ims-bordeaux.fr
Tél : 05 40 00 28 58 ; Fax : 05 56 37 15 45

Directeur adjoint : Sébastien PILLEMENT
PR2 : Ecole Polytechnique de l'Université de Nantes - IETR
sebastien.pillement@univ-nantes.fr
Tél : 02 40 68 30 64 ; Fax : 02 40 68 32 33

Sous-Directeur : Patrick GIRARD
DR1 CNRS : LIRMM – CNRS
patrick.girard@lirmm.fr
Tél : 04 67 41 86 29 ; Fax : 04 67 41 85 00

Mars 2021

Contributeurs

Lorena Anghel
Cécile Belleudy
François Berry
Lilian Bossuet
Emmanuel Boutillon
Daniel Chillet
Philippe Coussy
Dominique Dallet
Mickaël Dardaillon
Nathalie Deltimple
Patricia Desgreys
Sébastien Faucou
Abdoulaye Gamatié
Bertrand Granado
Patrick Girard
Guy Gogniat
Luc Hébrard
Jacques-Olivier Klein
Cristell Maneux
Kevin Martin
Ian O'Connor
Maxime Pelcat
Frédéric Pétrot
Sébastien Pillement
Andrea Pinna
Jean-Michel Portal
Olivier Romain
Gilles Sassatelli
Olivier Sentieys
Lionel Torres
François Verdier

SPINTEC /Grenoble INP
LEAT / Université Côte d'Azur
Institut Pascal / Université Clermont Auvergne
Lab. Hubert Curien / Univ. Jean Monnet
Lab-STICC / Univ. Bretagne Sud
IRISA / ENSSAT Lannion
Lab-STICC / Université Bretagne Sud
IMS / Bordeaux INP
IETR / INSA Rennes
IMS / ENSEIRB
LTCI / Telecom ParisTech
LS2N / Université de Nantes
LIRMM / CNRS
LIP6 / Sorbonne Université
LIRMM / CNRS
Lab-STICC / Université Bretagne Sud
ICube / Université de Strasbourg
C2N / Université Paris Sud
IMS / Université de Bordeaux
Lab-STICC / Université de Bretagne-Sud
INL / Ecole Centrale de Lyon
IETR / INSA Rennes
TIMA / Grenoble INP
IETR / Ecole Polytechnique de l'Université de Nantes
LIP6 / Sorbonne Université
IM2NP / Université Aix Marseille
ETIS / Cergy Paris Université
LIRMM / CNRS
IRISA / ENSSAT
LIRMM / Polytech Montpellier
LEAT / Université Côte d'Azur

lorena.anghel@phelma.grenoble-inp.fr
cecile.belleudy@unice.fr
francois.berry@univ-bpclermont.fr
lilian.bossuet@univ-st-etienne.fr
emmanuel.boutillon@univ-ubs.fr
daniel.chillet@irisa.fr
philippe.coussy@univ-ubs.fr
dominique.dallet@ims-bordeaux.fr
mickael.dardaillon@insa-rennes.fr
nathalie.deltimple@ims-bordeaux.fr
patricia.desgreys@telecom-paristech.fr
sebastien.faucou@univ-nantes.fr
abdoulaye.gamatie@lirmm.fr
bertrand.granado@sorbonne-universite.fr
patrick.girard@lirmm.fr
guy.gogniat@univ-ubs.fr
luc.hebrard@unistra.fr
jacques-olivier.klein@u-psud.fr
cristell.maneux@ims-bordeaux.fr
kevin.martin@univ-ubs.fr
ian.oconnor@ec-lyon.fr
mpelcat@insa-rennes.fr
frederic.petrot@imag.fr
sebastien.pillement@univ-nantes.fr
andrea.pinna@lip6.fr
jean-michel.portal@univ-amu.fr
olivier.romain@ensea.fr
sassatelli@lirmm.fr
olivier.sentieys@irisa.fr
lionel.torres@lirmm.fr
francois.verdier@univ-cotedazur.fr

Préambule

Ce rapport de prospective, rédigé à la demande du CNRS, décrit la vision du GdR SoC² sur les évolutions attendues dans la prochaine décennie des domaines couverts par le périmètre scientifique du GdR (system-on-chip, systèmes embarqués et objets connectés). Il s'agit d'un rapport stratégique et important permettant d'effectuer une analyse du domaine et de formuler des recommandations pour la politique scientifique du CNRS (postes, instruments, soutien, articulation avec l'Europe ...)

Ce rapport est le fruit de réflexions initiées au sein du GdR en juillet 2020. Un processus de sondage a permis aux responsables d'axe d'identifier des thématiques : celles en émergence, celles qui sont existantes et fortes, et celles ne présentant plus de véritables verrous scientifiques.

Ces visions ont été présentées en novembre 2020 devant les comités stratégique et d'animation du GdR SoC² ainsi que devant des personnalités externes (représentants d'acteurs industriels / proche industrie du domaine dans le paysage national - Thales, Safran, CEA, INRIA ...), permettant de recueillir des retours et d'enrichir ainsi l'analyse prospective.

La phase finale de rédaction du rapport et des recommandations s'est appuyée sur les analyses et les échanges avec les comités et les personnalités externes. Nous avons à cœur de pointer les forces et les faiblesses de la recherche académique dans notre domaine en France, et d'associer à cette analyse des recommandations pour renforcer, soutenir, rendre plus efficaces nos recherches, aussi bien en France que dans le contexte du nouveau programme cadre Horizon Europe.

Ce rapport est structuré en trois parties :

- une courte présentation du GdR et de ses missions pour situer le contexte du rapport
- une analyse prospective des domaines scientifiques couverts par le GdR à l'horizon 2030
- un résumé des principales conclusions et des recommandations

CONTRIBUTEURS.....	2
PREAMBULE.....	3
PRESENTATION DU GDR	5
1. OBJECTIFS DU GDR	6
2. POSITIONNEMENT DU GDR	7
ANALYSE PROSPECTIVE	9
1. CALCUL EMBARQUE HAUTE PERFORMANCE	11
2. FRONTIERES ET INTERFACES CYBER-PHYSIQUES	14
3. SECURITE ET INTEGRITE DES SYSTEMES	17
4. OBJETS CONNECTES	21
5. TECHNOLOGIES DU FUTUR.....	26
6. METHODES ET OUTILS.....	29
7. IA ET SYSTEMES EMBARQUES	33
SYNTHESE DES CONCLUSIONS ET RECOMMANDATIONS.....	37
1. IDENTIFICATION DES DEFIS ET LES PISTES DE RECHERCHE A PRIVILEGIER.....	38
a. <i>Des contextes à forts enjeux – évolutions attendues et pistes de recherche.....</i>	38
i. <i>Technology push – Scaling and complexity.....</i>	38
ii. <i>Technology push – hétérogénéité.....</i>	40
iii. <i>Application pull – Internet of Everything (IoE).....</i>	43
iv. <i>Matériel open-source</i>	45
v. <i>Consommation énergétique ICT et développement durable.....</i>	46
vi. <i>Souveraineté européenne en microélectronique</i>	47
vii. <i>Intégrité et fiabilité.....</i>	49
viii. <i>Sécurité et confidentialité.....</i>	50
b. <i>Des enjeux à fortes valeurs stratégiques : les atouts de la France et de l'Europe.....</i>	51
i. <i>Communications – vers la 6G et les THz.....</i>	51
ii. <i>Mémoire : nouvelles technologies, nouveaux usages, nouveaux paradigmes de calcul</i>	51
iii. <i>Intelligence artificielle embarquée</i>	53
iv. <i>Calcul quantique.....</i>	54
2. RECOMMANDATIONS STRUCTURELLES – INSTRUMENTS A PRIVILEGIER POUR DYNAMISER LA RECHERCHE	55
a. <i>Articulation formation – recherche – industrie.....</i>	55
b. <i>Favoriser la pluridisciplinarité.....</i>	56
c. <i>Passage à l'échelle – une plateforme pérenne multi-partenaire / multi-institut.....</i>	56
d. <i>Moyens et profils de poste à privilégier.....</i>	57
e. <i>Des actions phares pluridisciplinaires : IA embarquée et calcul quantique</i>	58

GdR SoC²

System On Chip, Systèmes embarqués et Objets Connectés

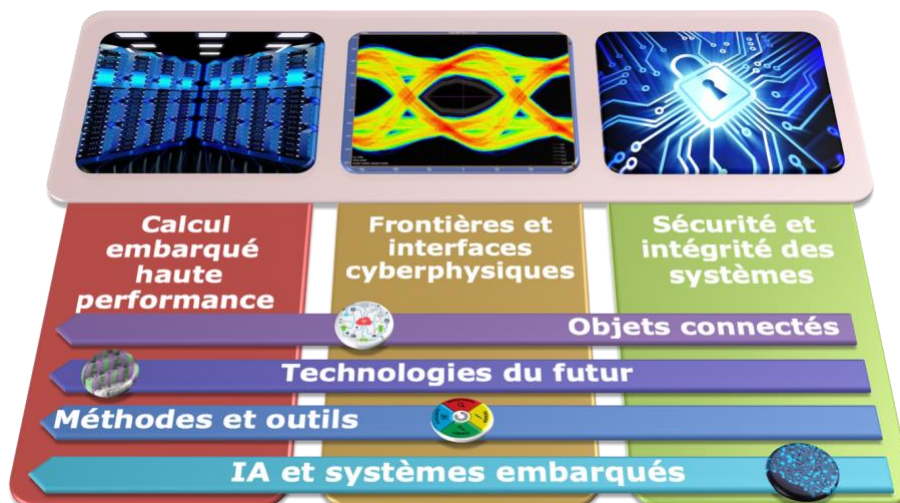


Présentation du GdR

1. Objectifs du GdR

L'objectif de ce Groupement de Recherche est d'étudier et de proposer de nouvelles approches pour la conception et la validation des systèmes embarqués pour les objets connectés, le "*edge computing*" et l'intelligence artificielle embarquée. L'accent est mis sur les architectures matérielles en prenant en compte leurs interactions avec les logiciels (applications, systèmes d'exploitation, reconfiguration), leurs outils de développement et leur environnement réel (analogique, communication radiofréquence et haut-débit, capteurs et actuateurs). Des considérations technologiques de la micro-électronique complète les domaines couverts par le GdR. Dans ce cadre, les principaux défis actuels sont nombreux (liste non exhaustive) – réduire la consommation énergétique pour l'autonomie des systèmes embarqués et pour maîtriser le bilan carbone du calcul exascale, garantir la sécurité et l'intégrité des systèmes électroniques, maîtriser les coûts de conception et de validation des systèmes embarqués et de calcul, assurer l'adéquation des systèmes intégrés dans les objets connectés pour de multiples secteurs d'application (territoire intelligent, e-santé, sécurité des biens et des personnes, usine 4.0 ...). S'appuyant sur des systèmes intégrés complexes intégrant plusieurs milliards de dispositifs élémentaires sur une puce de silicium (SoC) et utilisant de nouveaux paradigmes de calcul, de nouvelles technologies ou approches intégratives (en trois dimensions par exemple), ces systèmes requièrent un large panel de compétences pluridisciplinaires en électronique (architecture et méthodologie) et micro-nanoélectronique (logique, analogique, RF), en informatique embarquée et temps-réel (système d'exploitation), et en physique (technologies émergentes).

Le GdR SoC² propose de traiter ces différents défis de manière holistique et transversale en rassemblant différentes communautés impliquées dans ces problématiques. Il est structuré en 7 axes – 3 axes thématiques (Calcul embarqué haute performance, Frontières et interfaces cyberphysiques, Sécurité et intégrité des systèmes) et 4 axes transverses (Objets connectés, Technologies du futur, Méthodes et outils, IA et systèmes embarqués) :



Le GdR SoC² regroupe 67 laboratoires en France (dont 19 ayant adhéré depuis le renouvellement en 2018), principalement des UMR CNRS, 4 organismes externes, et plus de 800 permanents (membres d'UMR ou EA INSIS et INS2I - chercheurs CNRS sections 7 et 8, enseignants-chercheurs sections 27, 61 et 63 du CNU, enseignants-chercheurs rattachés à des écoles d'ingénieur privées, membres des EPI-INRIA).

2. Positionnement du GdR

Le domaine de la conception des systèmes sur puce, systèmes embarqués et objets connectés est un domaine clé pour la Société de l'Information et de la Communication dans laquelle nous vivons, et plus particulièrement pour insuffler, accompagner et soutenir la mutation numérique. Les frontières du domaine s'étendent de la recherche sur les nouveaux composants micro-nano-électroniques / photoniques, à la recherche sur le calcul avancé et la gestion de l'information et de l'énergie, liée aux enjeux de l'informatique pervasive. De plus, il requiert de la continuité entre ces grandes disciplines, intégrant les nouveaux dispositifs dans de nouvelles architectures et de nouveaux paradigmes de calcul (pour le calcul haute-performance et pour les systèmes embarqués où la consommation énergétique est critique), de nouveaux moyens de stockage (comme les mémoires non-volatiles), de nouveaux médias de communication (RF et optique) et d'interaction (capteurs et actuateurs). Les recherches dans le domaine des systèmes sur puce, fortement liés à l'évolution des technologies d'intégration, ont ouvert la voie à la création et à l'essor d'objets devenus maintenant indispensables : les cartes à puce, les smartphones et les tablettes (convertisseurs de données, transceivers RF CMOS, imageurs haute-résolution...), l'accès internet très haut-débit (ADSL2+, FTTH, LTE / 5G), ou encore les centres de calcul haute performance et les centres de traitement de données (architectures mémoires, communication très haut-débit, parallélisme massif...), qui sont au cœur de la numérisation de la société. Les travaux sur les systèmes sur puce sont aussi en interaction forte avec le monde socio-économique (STMicroelectronics, Thales, NXP, Valeo, Bull, Airbus Defense & Space, etc.) sans oublier les nombreuses PME/ETI de ce secteur économique présentes en France et en Europe.

Ce GdR est par essence au cœur même de cette révolution numérique que nous vivons tous les jours. Naturellement, il s'articule avec les défis des composants et du calcul avancé, mais également avec (i) l'internet du futur et des objets en ce qui concerne l'infrastructure de communication, (ii) la robotique en ce qui concerne les systèmes cyberphysiques, (iii) la cybersécurité en ce qui concerne les techniques matérielles-logicielles pour la sécurité et la robustesse des systèmes intégrés.

Le domaine couvert par le GdR est fortement interdisciplinaire entre les aspects matériels et logiciels. Le GdR SoC² occupe ainsi une position originale à l'interface entre les communautés électronique, physique et informatique, et donc à l'interface entre les Instituts INSIS, qui est son rattachement principal, et INS2I du CNRS. Au CoNRS, ses membres sont principalement rattachés aux sections 7¹ et 8². De plus, de nombreux membres du GdR SoC² sont des enseignant-chercheurs, appartenant aux sections 27³, 61⁴ et 63⁵ du CNU, lorsqu'ils sont dans des établissements publics (quelques membres étant rattachés à des écoles d'ingénieur privées).

¹ Sciences de l'information : signaux, images, langues, automatique, robotique, interactions, systèmes intégrés matériel-logiciel

² Micro et nanotechnologies, électronique, photonique, électromagnétisme, énergie électrique

³ Génie Informatique (temps réel, architecture, réseaux), Automatique (automatique continue, robotique, productique), Traitement du Signal (information, signal, image, vision)

⁴ Algorithmique et combinatoire, Architecture des machines et des systèmes, Bio-informatique, Communication homme-machine, Génie logiciel et programmation, Informatique théorique ou fondamentale, Imagerie numérique, Intelligence artificielle, Recherche opérationnelle, Réseaux, Systèmes d'information, Systèmes informatiques

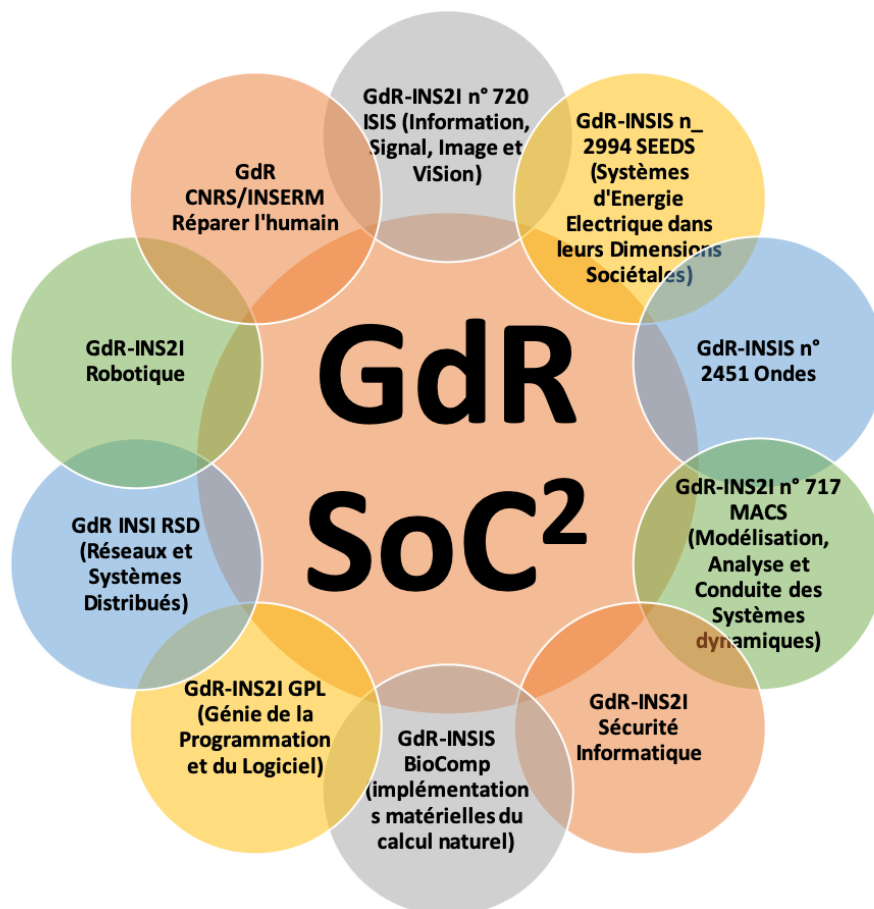
⁵ Composants et systèmes électriques, systèmes et composants électroniques et optiques, micro et nanoélectronique, matériaux pour l'électronique et le génie électrique, électronique organique, photonique, télécommunications, antennes, électromagnétisme, micro-ondes, traitement du signal appliqué, acoustique, éclairage, instrumentation, génie biomédical, électrotechnique, systèmes d'éclairage, commande des systèmes électriques, électronique de puissance, plasmas froids, plasmas thermiques

Le GdR rassemble une large communauté intégrée dans de nombreuses UMR relevant des instituts CNRS INSIS et INS2I, mais aussi dans des équipes d'accueil (EA) et des EPI-INRIA. Des liens forts sont établis avec d'autres EPST et EPIC tels que le CEA, et avec l'industrie.

Pour la communauté SoC², ce GdR constitue l'unique structure nationale d'animation de la recherche. Le GdR a de nombreuses interactions avec d'autres organismes à l'échelle nationale dont notamment:

- le Centre National de Formation en Microélectronique (CNFM, créé en 1981, pour la formation et l'enseignement) – avec la collaboration de Thales, le GdR SoC² a participé au montage et à la mise en place du concours national "[RISC-V student contest](#)"
- l'Unité Mixte de Service "[Circuits Multi-Projets](#)" (UMS 3040 CMP, créée en 1981) qui fournit principalement un accès aux technologies de fabrication de circuits intégrés
- [Embedded France](#) (association industrielle dans le domaine des systèmes embarqués), dont le GdR SoC² est adhérent
- le [Comité Stratégique de Filière \(CSF\) Électronique](#), dont le GdR SoC² est partie prenante

Sur la carte des GdRs (figure ci-dessous), les GdRs connexes relèvent d'INSIS (SEEDS, Ondes, BioComp) et d'INS2I (ISIS, RSD, Robotique, MACS, Recherche Opérationnelle, GPL, Sécurité). Il est également à souligner que de nombreux membres du GdR SoC² sont également impliqués dans (voire étaient associés au montage) d'autres GdR ; la force du GdR SoC² est de s'appuyer sur son spectre large pour assurer les connexions entre les disciplines et tisser des collaborations fortes avec de nombreuses communautés sur des thématiques connexes.



GdR SoC²

System On Chip, Systèmes embarqués et Objets Connectés



Analyse Prospective

Dans les paragraphes qui suivent, une description de chaque axe est donnée (contenu et périmètre scientifique de l'axe) ainsi qu'une analyse prospective horizon 2030.

Pour chaque thématique, nous identifions :

- L'évolution attendue du domaine dans les 10 prochaines années (horizon 2030),
- Les défis importants et/ou émergents pour le domaine,
- Les verrous liés aux défis et les objectifs scientifiques,
- Les forces et les faiblesses de la recherche scientifique française.

1. Calcul embarqué haute performance

Mots clés

Calcul : embarqué, haute performance, parallèle, temps réel, prédictible, sûr

Logiciel embarqué : compilation, adéquation algorithme-architecture, runtime, gestion de ressources, hyperviseur, modèles de programmation

Architecture : hétérogénéité, reconfiguration, autonomie, adaptabilité, microarchitecture, interconnexions, accélérateurs, MPSoC⁶, multi-cœur, efficacité énergétique

Le périmètre de l'axe Calcul Embarqué Haute Performance couvre les modèles et les méthodes de calcul, ainsi que l'exploitation des systèmes numériques, principalement embarqués, en considérant les aspects logiciels et matériels. Les travaux ont pour but de concevoir des systèmes durables, efficaces, prédictibles, et en capacité d'exécuter des applications intégrant des algorithmes complexes, contribuant ainsi à la construction d'une société numérique ouverte et citoyenne.

Thèmes actuels et évolutions attendues à l'horizon 2030

D'ici 2030, la recherche scientifique associée à l'axe Calcul Embarqué Haute Performance va devoir prendre en considération un ensemble d'évolutions déjà visibles.

Tout d'abord, une **évolution des usages**, qui fait écho aux enjeux sociétaux stratégiques, en particulier le développement durable et la création d'une société numérique ouverte et citoyenne. Concernant le *développement durable*, il s'agit de minimiser l'impact environnemental des systèmes. Les compétences présentes au sein de l'axe permettront de travailler sur les phases de conception et d'exploitation des systèmes numériques. Les travaux sur l'autonomie et l'efficacité énergétique des systèmes doivent être poursuivis. Des travaux émergents devront être soutenus, concernant la durabilité et la réutilisabilité accrue des systèmes (favorisées par exemple par l'utilisation de composants matériels et logiciels libres), ou l'intégration de capacités d'adaptation. La création d'une *société numérique ouverte et citoyenne* nécessitera de savoir construire des systèmes respectueux de la vie privée, à même d'être inscrits dans des plans de gouvernances des données. Plus largement, il s'agira de prendre en considération les aspects éthiques des applications cibles, en intégrant en particulier une réflexion sur la notion de sobriété numérique, faisant ainsi le lien avec le développement durable.

Ensuite, une **évolution des paradigmes de calcul**. Les divers types de traitements de données en périphérie (*edge computing*, *near-sensor computing*) apparaissent d'ores et déjà comme des solutions incontournables : en contribuant à améliorer l'efficacité énergétique (via la réduction du volume des communications), en offrant un meilleur respect de la vie privée (via un traitement local des données brutes), et en promettant des latences réduites pour la réalisation des services. Cette évolution s'accompagne de problématiques inédites en termes de capacité de calcul embarquée, sous contrainte de ressources (principalement énergie, mais aussi mémoire), et maintenant de prédictibilité pour répondre au besoin des fonctions temps-réel. Ces tendances ne seront qu'accentuées par la généralisation de l'IA embarquée. Des modèles de calcul dédiés seront nécessaires pour atteindre ces objectifs. Les travaux qui visent à définir de tels modèles devront donc être poursuivis et soutenus. On peut citer, à titre d'illustration et de façon non exhaustive, le calcul approximé, le calcul intermittent, le calcul en mémoire, ou encore le calcul neuromorphique.

Enfin, une **évolution des architectures et des technologies**. Aujourd'hui, l'exécution efficace des algorithmes issus de l'IA par exemple, que ce soit en périphérie ou sur le *cloud*, passe par l'intégration

⁶ Multi-Processor System on Chip

d'accélérateurs au sein des architectures. Il peut s'agir de (GP)GPU⁷, de TPU⁸, de solutions dédiées à base de FPGA⁹, et sans doute, à terme, d'accélérateurs exotiques, y compris quantiques. Par ailleurs, le caractère prometteur des nouvelles technologies mémoires (par exemple STT/SOT-MRAM¹⁰, PCM¹¹, ReRAM¹²), et d'interconnexion des composants (par exemple des interconnexions optiques ou l'exploitation de l'empilement 3D – *3D-stacking*) permet d'envisager une remise en cause plus large des architectures des systèmes, au-delà d'une "simple" intégration d'accélérateurs. Une telle révolution devra entre autres permettre d'améliorer fortement l'efficacité énergétique du calcul, qui est pénalisée aujourd'hui par le coût du mouvement des données entre circuits mémoire et circuits de calcul. Des architectures innovantes capables de réaliser efficacement du calcul embarqué haute performance, en tirant bénéfice des avantages des solutions évoquées précédemment, devront être recherchées.

Un exercice de synthèse de ces différentes évolutions nous amène à formuler deux grands objectifs scientifiques pour 2030 :

- **Maîtriser l'hétérogénéité des systèmes**, à tous les niveaux : technologique, architectural, et logiciel. Cet objectif, à visée clairement multi-niveau, joue un rôle absolument central dans la définition de solutions répondant aux exigences induites par les évolutions décrites précédemment. En effet, grâce aux leviers spécifiques offerts par chaque paradigme de calcul, architecture et technologie, diverses problématiques en jeu pour l'avènement de systèmes à la fois performants, efficaces en énergie et prédictibles seront plus aisément adressables.
- **Favoriser le développement des systèmes écoresponsables**, en visant en particulier la durabilité et la réutilisation des systèmes et/ou de leurs composants. Cet objectif pourrait être considéré comme transversal au GdR SoC². Les aspects logiciels et matériels visés par le présent axe sont naturellement au centre des réflexions déjà entamées au sein de la communauté scientifique nationale et internationale.

Analyse des thématiques - défis importants et/ou émergents

Au regard des thématiques et des évolutions pressenties dans cet axe, la recherche de futures solutions candidates passera nécessairement par la prise en compte d'un certain nombre de défis clés, pouvant être déclinés selon les deux objectifs retenus ci-dessus :

- **Défi technologique** : il s'agit ici d'approprier les paradigmes de calcul et les technologies identifiés comme étant incontournables ou prometteurs (comme le calcul quantique par exemple). Cela facilitera des choix judicieux de briques de base selon les exigences/besoins attendus des systèmes de demain.
- **Défi méthodologique pour la conception** : une fois des ingrédients adaptés sélectionnés, il faudra maîtriser leur intégration/combinaison de manière à satisfaire les exigences/besoins à l'échelle des systèmes cibles. Il s'agit de sélectionner et combiner des méthodes développées par ailleurs pour arriver à une méthodologie de conception pour l'architecture complète d'un système.
- **Défi sociétal** : il faudra anticiper des réponses aux attentes sociétales (société numérique ouverte et citoyenne), en tenant compte des problématiques associées telles que l'écoresponsabilité,

⁷ (General Purpose) Graphics Processing Unit

⁸ Tensor Processing Unit

⁹ Field Programmable Gate Array

¹⁰ Spin Transfer Torque / Spin Orbit Torque Magnetic Random Access Memory

¹¹ Phase Change Memory

¹² Resistive Random Access Memory

l'éthique, l'interopérabilité ou la souveraineté numérique, dans les choix et la mise en œuvre des mécanismes caractérisant les systèmes.

On notera que les trois défis précédents sont en prise avec des préoccupations relevant des autres axes du GdR SoC².

Forces et faiblesses de la recherche scientifique française

Forces - La recherche à l'interface matériel/logiciel (par exemple compilation, modèles de calcul, adéquation algorithme architecture, temps réel, etc.) - Communautés actives et dynamiques autour du logiciel libre et du matériel libre - Compétences fortes et reconnues sur la conception des systèmes critiques - De nouveaux besoins apparaissent (IA embarquée)	Faiblesses - Hormis le CEA, il manque d'équipes ayant les moyens humains et technologiques pour entreprendre des projets d'envergure en architecture (type ETHZ par exemple) - Très peu d'équipes spécialisées en microarchitecture, faible soutien au niveau national sur ce thème
Opportunités Puisque l'industrie française comprend des acteurs majeurs (au moins à l'échelle européenne) sur les aspects architecturaux et technologiques, un partenariat pérenne plus resserré avec les acteurs académiques, avec les moyens adéquats, serait à encourager. En effet, les deux premiers défis ne pourront être adressés convenablement que si les détenteurs de technologies (i.e. industriels) et académiques travaillent main dans la main	Menaces - L'Europe est souvent perçue comme étant "à la traîne" pour cette thématique - Du fait de cette "érosion thématique des équipes", les aspects (micro)architectures pourraient à terme connaître une perte de vitesse - La pyramide des âges sur cette thématique ne semble pas favorable

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : C2N, CITI, CRISTAL, ETIS, Heudiasyc, I3S, I-Cube, IETR, IMS, Institut Pascal, IRISA, IRIT, LAAS, LaBRI, Lab-STICC, LAMIH, LE2I, LEAT, LIG, LIP, LIP6, LIGM, LIRMM, LS2N, LTCI, TIMA.

Autres acteurs institutionnels : CEA, Inria, ONERA

2. Frontières et interfaces cyber-physiques

Mots clés

Intelligence embarquée : systèmes AMS¹³ intelligents, IA embarquée

Communications : communications RF¹⁴, modularité/agilité

Interfaçage : systèmes hétérogènes, conversion d'énergie embarquée

Les mondes informatique et physique fusionnent pour évoluer vers les systèmes cyber-physiques. Le monde physique entre dans le réseau avec une interaction accrue entre les informations physiques ultra-riches du monde réel et la haute puissance de calcul du monde numérique. Cette fusion vise à accroître la capacité d'adaptation, l'autonomie, l'efficacité, la fonctionnalité, la fiabilité et la sécurité des systèmes du futur.

L'interfaçage du système d'objets connectés d'un côté avec le monde physique via des capteurs et actuateurs et de l'autre côté avec le cœur du système via des communications, notamment RF, nécessite des recherches pour relever les défis des systèmes cyber-physiques intelligents, et notamment pour (i) intégrer une intelligence localisée au niveau de l'objet ; (ii) optimiser la communication avec le cœur du système ; et (iii) développer des modèles de haut-niveau de ces nœuds sensibles.

Thèmes actuels et évolutions attendues à l'horizon 2030

Les thèmes de recherche récents se sont donc concentrés sur :

- La nécessité d'intégrer une intelligence localisée au niveau de l'objet (nœud sensible), et le fait que cette intelligence doit consommer le moins possible, en accord avec l'application. Pour cela les verrous adressés sont notamment :
 - L'optimisation de l'élément de transduction, d'acquisition et de numérisation de l'information, soit au niveau même de l'élément, soit en ajoutant une électronique spécifique permettant d'optimiser sa consommation tout en assurant une haute résolution de mesure
 - L'efficacité énergétique de la transmission RF et une gestion flexible des phases d'activité/d'inactivité du capteur pour permettre une réduction de la consommation.
 - La compression des données localement (Analog to Information) par le développement d'algorithme spécifique, adaptés à l'application : numérisation et transmission d'un faible nombre d'échantillons en deçà de la limite de Nyquist, suffisant pour calculer l'information utile pour l'application. Le pré traitement analogique et la cadence d'échantillonnage doivent être flexibles.
- La nécessité d'optimiser la communication avec le cœur du système. Ici, les verrous portent sur :
 - Le développement de communications RF interopérables
 - Des communications compatibles avec la sécurité des données, et les techniques développées dans l'axe "sécurité et intégrité des systèmes"
 - L'optimisation de la consommation de la communication
- La nécessité de développer des modèles de haut niveau de ces nœuds sensibles, compatibles avec les techniques de simulation/modélisation de systèmes développées dans l'axe "Méthodes et outils de conception, simulation, évaluation et vérification des systèmes et systèmes de systèmes". Ici, le principal verrou est lié à la précision du modèle qui se doit d'être suffisante en

¹³ Analog and Mixed-Signal

¹⁴ Radio Frequency

tenant compte des principaux paramètres influant sur le système, tout en permettant une simulation système à grande échelle.

Le contexte d'évolution de ces thématiques pour les dix prochaines années est

- Une évolution qui sera pilotée de plus en plus par les applications et les systèmes plutôt que par les technologies. A ce titre, l'IRDS (International Roadmap for Devices & Systems) s'est substituée à l'ITRS (International Roadmap for Technology for Semiconductors), car on attend de forts gains sur l'ensemble des défis grâce au déploiement des Systèmes Connectés (SC) à grande échelle et leur utilisation à l'échelle de grands systèmes (industrie 4.0).

Les autres grands domaines d'applications qui poussent l'innovation, dans les technologies et la conception de circuits sont la santé, la sécurité, l'énergie/ l'environnement et les transports.

Par exemple, la santé numérique requiert la miniaturisation des dispositifs biomédicaux et l'intégration de systèmes intégrés intelligents. L'interfaçage de composants et de systèmes électroniques avec des systèmes biologiques offrira une connexion transparente au corps pour une surveillance continue, ainsi qu'à des fins d'électrostimulation.

- Le développement des systèmes connectés et des capacités de l'intelligence artificielle (IA) s'est considérablement accéléré ces dernières années et va poursuivre cette ascension ; cette avancée s'accompagne d'une montée en puissance constante et nécessaire du "numérique". La prochaine étape pour l'IA sera de communiquer de façon très interactive avec le monde physique, c.à.d. avec l'humanité, le monde du vivant et la matière. Nous sommes dans l'ère de l'IA embarquée. Il s'agit à présent d'utiliser ces nouvelles capacités pour construire des systèmes intelligents hautement sophistiqués basés sur le matériel et les logiciels qui garantissent un traitement, un stockage et un transfert de données efficaces, sûrs, sécurisés et durables.

- Réduire la consommation d'énergie/ développement durable :

Pour rendre réalisable les systèmes complexes : le principal verrou identifié est l'excès de consommation des systèmes analogiques actuels qui empêche un développement rapide de l'IA sur les réseaux de capteurs et d'actuateurs ainsi que les réseaux de communications radiofréquence ou système communicants autonomes.

La vision pour 2030 est d'atteindre l'objectif politique actuel de l'UE de 30% d'économies en utilisant des solutions innovantes basées sur les composants électroniques et les systèmes.

Pour le développement durable et l'énergie, face à l'explosion du nombre de composants et à la consommation de matériaux rares, une recherche est attendue pour la durabilité des objets.

Analyse des thématiques - défis importants et/ou émergents

Ainsi dans les prochaines années, nous savons qu'il y aura toujours plus de systèmes connectés et l'évolution sera pilotée par les systèmes plutôt que par les technologies.

Les SC vont se développer spécifiquement par application et selon les usages induits. Toutefois, au niveau de l'interfaçage avec le monde physique d'une part, et avec l'intelligence du système d'autre part, des défis communs à tous les SC se dégagent.

- Nécessité d'intégrer une intelligence électronique localisée (IA embarquée) au niveau de l'objet (nœud sensible) pour optimiser les performances :
 - Adaptabilité à l'application et à l'environnement de déploiement, prise en compte du packaging
 - Compatibilité matériaux : bio compatibilité, communication intracorporelle.
 - Extraction de l'information pertinente localement (Analog to Feature Converter) par le développement d'algorithme spécifique, adaptés à l'application : numérisation et transmission d'un faible nombre d'échantillons.
 - Utilisation de l'IA pour optimiser les paramètres de conception des circuits hétérogènes

- Implémentation analogique de l'IA
- Nécessité d'adapter la communication avec le cœur du système à l'application.
 - Conception de communications RF pour les réseaux 5G-6G large bande, flexibles et interopérables.
 - Conception de communications RF pour l'IoT très efficace énergétiquement : communications durables, IoT coopératif.
 - Gestion intelligente et flexible des phases d'activité/d'inactivité de la transmission RF pour une réduction drastique de la consommation.
- Nécessité de développer des modèles de haut-niveau de ces nœuds sensibles
 - Compatibles avec les techniques de simulation/modélisation de systèmes hétérogènes (i.e. intégrant des sous-systèmes de natures différentes) développées dans l'axe "Méthodes et outils"
 - Avec une précision du modèle suffisante en tenant compte des principaux paramètres influant sur le système, tout en permettant une simulation système à grande échelle

Forces et faiblesses de la recherche scientifique française

Forces ● Beaucoup d'acteurs excellents dans les domaines des capteurs, des systèmes hétérogènes et de la RF, ainsi qu'en modélisation système. ● L'IA est un domaine plus récent mais largement investi par les laboratoires français. ● Les formations des futurs doctorants intègrent déjà tous ces concepts. Nous sommes donc bien placés et aguerris pour relever les futurs défis	Faiblesses ● L'innovation se passe à l'interface entre plusieurs domaines. Les jeunes chercheurs capables de maîtriser à la fois la conception de circuits et les algorithmes sont encore peu nombreux, des recrutements sont à prévoir sur ce type de profil. ● Forte captation des étudiants (Master-ingénieurs) / doctorants formés sur ces sujets par l'industrie qui limite les recrutements académiques.
Opportunités ● Expertise nationale et européenne reconnue dans le domaine du "More than Moore" ● Engouement pour l'IoE¹⁵, interactions fortes entre le monde physique et le monde numérique ● Stabilité des technologies clés, maîtrise de l'intégration de divers types d'éléments hétérogènes	Menaces ● L'importance du matériel dans les systèmes embarqués fait que la dynamique industrielle est très significative sur plusieurs sujets relatifs, il est crucial que les liens entre les industriels et la communauté académique soient maintenus voire renforcés.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : C2N, ETIS, FEMTO-ST, GeePs, ICube, IEMN, IM2NP, IMS, INL, LAAS, Lab-STICC, LCIS, LEAT, LIP6, LIRMM
 IMT-TP : LTCI

¹⁵ Internet of Everything

3. Sécurité et intégrité des systèmes

Mots clés

Sécurité matérielle

Test, conception en vue du test, test intégré, simulation de fautes, diagnostic

Fiabilité, robustesse, tolérance aux fautes

Vérification

L'environnement des SoC² évolue vers plus de complexité, de connectivité et de mobilité. Le maître mot de leur développement futur est la confiance : la capacité des systèmes à résister à des attaques contre la confidentialité et l'intégrité des informations, la traçabilité fournie par les systèmes quant à leur conception et à leur origine (informations nécessaires pour combattre la contrefaçon et la fraude), la possibilité d'identifier un dysfonctionnement lié aux processus de conception ou de fabrication et d'en comprendre les origines, le bon fonctionnement des systèmes soumis à des perturbations environnementales, et la robustesse des systèmes face aux variations de paramètres qui peuvent apparaître lors de leur fabrication et/ou pendant leur durée de vie.

La confiance matérielle est passée depuis quelques années au premier plan et pose plusieurs défis dont (i) l'adaptation des méthodologies aux systèmes ultra-faible-consommation, systèmes multi-horloge, circuits mixtes et technologies avancées (intégration 3D, nanoélectronique) et du futur (spintronique, nanotubes de carbone, électronique souple ...) ; (ii) la montée dans les niveaux d'abstraction pour l'insertion des méthodes développées (vérification formelle, conception en vue du test, tolérance aux fautes, sécurité) dans les flots de conception ; et (iii) le passage à l'échelle en termes de complexité, densité d'intégration, de fusion du logiciel et du matériel, et d'hétérogénéité des systèmes.

Thèmes actuels et évolutions attendues à l'horizon 2030

La cybersécurité des systèmes électroniques embarqués et des objets connectés est un enjeu majeur actuel autant pour les aspects sociétaux que pour les aspects économiques. Il convient d'apporter des solutions pour **sécuriser les systèmes électroniques embarqués** des secteurs névralgiques (hôpitaux et santé, énergie, transports, industrie 4.0, ...) et d'améliorer l'évaluation de la sécurité et de la fiabilité de ces systèmes. De plus, le matériel doit permettre d'assurer une garantie de bas niveau concernant la sécurité des traitements de données dans des environnements qui ne sont pas sûrs. L'enjeu tient donc dans la capacité de la communauté à adresser des chemins d'attaques complexes (dont la complexité évolue avec celle des cibles) exploitant conjointement des failles cryptographiques, logicielles, matérielles et physiques, de proposer des contre-mesures efficaces et d'en évaluer l'efficacité.

La résilience des systèmes électroniques embarqués et des objets connectés est un enjeu spécifique qui doit tendre vers des **capacités d'autodiagnostic et d'autoréparation des systèmes**. Que ce soit face à des actes malveillants (cyberattaques) ou des perturbations environnementales, le système doit être en mesure de retrouver au plus vite un état de fonctionnement normal et garanti. Les tests embarqués, le diagnostic, la fiabilité et la sécurité doivent être pensées dans cette perspective. Dans ces domaines de nouveaux paradigmes doivent être étudiés et développés pour atteindre ces objectifs.

D'ici 2030 de nouvelles technologies seront étudiées que ce soit pour le traitement (*near- ou in-memory computing, approximate computing*, etc.) et la mémorisation (mémoires magnétiques) des données, l'intégration multidimensionnelle des puces électroniques et les connexions internes à haute densité. **Le test, la fiabilité et la sécurité** (évaluation vis-à-vis des attaques connues,

exploitation pour des applications de sécurité) doivent être considérés avec la plus grande attention pour ces **nouvelles technologies** pour accompagner leur développement et assurer leur pérennité industrielle.

D'ici 2030, le **déploiement de l'utilisation de l'IA** dans toutes les thématiques concernées par cet axe ne sera pas finalisé. Il existe encore de nombreuses possibilités de déploiement de l'IA dans nos thématiques, mais les chercheurs doivent considérer la compréhension des mécanismes de l'IA pour leur domaine comme une nécessité absolue afin de tendre vers la meilleure utilisation possible de ces techniques mais aussi afin d'en déterminer précisément les limites.

Analyse des thématiques - défis importants et/ou émergents

Thématiques en émergence et à soutenir : architectures de processeurs sécurisés par conception, exploitation et spécialisation des techniques d'apprentissage automatique pour l'amélioration des attaques physiques, sécurité des SoC complexes et hétérogènes ; intégrité des processeurs neuro-inspirés, des circuits approximatifs, et des architectures de calcul en mémoire.

- La recherche d'architectures de processeurs sécurisés par conception contre les attaques logicielles et physiques qu'elles soient par injection de fautes ou par l'analyse de canaux cachés est un enjeu très fort actuellement et pour lequel la concurrence est déjà rude. Un effort de recherche considérable est consacré à l'évolution des processeurs libres (comme ceux basés sur l'ISA¹⁶ RISC-V) pour proposer des architectures plus sécurisées.
- L'utilisation, la spécialisation et la compréhension des outils et méthodes de Machine Learning pour améliorer les attaques par analyse de canaux cachés est une thématique en pleine émergence avec déjà de très belles contributions de la communauté française au meilleur niveau international. Des recherches approfondies sur la proposition de contre-mesures spécifiques à ces nouvelles attaques doivent être menées. Vu l'ampleur de l'utilisation des algorithmes du domaine de l'IA, la fiabilité, le test et la résistance vis-à-vis des attaques physiques de leurs implémentations matérielles est également un enjeu crucial de la prochaine décennie.
- La sécurité interne aux SoC hétérogènes et complexes est aussi une thématique en émergence, de nouveaux chemins d'attaques (souvent mixte ou à la frontière entre le logiciel et le matériel) permettent de mettre en œuvre des attaques par analyse, par injection de fautes et par compromission d'une partie du SoC par une autre partie du SoC. Ces chemins d'attaques doivent être étudiés plus en profondeur et des protections dédiées doivent être proposées.
- Les plateformes d'injection de fautes sont de plus en plus perfectionnées et abordables, de nouveaux canaux sont explorés et permettent d'envisager de nouveaux chemins d'attaque.
- On trouve aujourd'hui des processeurs neuro-inspirés utilisant l'apprentissage profond (réseaux de neurones) pour des applications spécifiques (reconnaissance, classification, recherche de données) dans les domaines de la médecine, de la robotique, ou de la sécurité des personnes. Dans ce contexte, l'intégrité de ces processeurs doit être garantie par des techniques de test et de fiabilité spécifiques et fiables.
- Dans le cas des circuits utilisés pour de nouveaux paradigmes de calcul (comme le calcul approximatif), la phase de test doit être repensée et reconsidérée. En effet, lorsqu'un circuit est déclaré fautif, il est important de déterminer si les erreurs observées peuvent être malgré tout acceptées ou pas (perte de qualité acceptable selon un seuil fixé au préalable). De nouvelles stratégies de test sont donc attendues dans ce domaine.
- Comme tous les systèmes électroniques intégrés, les processeurs utilisant des architectures de calcul en mémoire doivent être testées à l'issue de leur fabrication. Pour autant, à ce jour, il

¹⁶ Instruction Set Architecture

n'existe pas de techniques spécifiques dédiées à ce type d'architecture. Les travaux réalisés dans le domaine du test des mémoires pourront servir de base au développement de solutions spécifiques aux architectures de calcul en mémoire.

Thématiques existantes fortes qu'il convient de maintenir ou de renforcer : nouveaux moyens pour les attaques physiques par injection de fautes, contre-mesures technologiques aux attaques physiques, implémentation (logicielle et matérielle) efficace d'algorithmes de cryptographie, outil de CAO pour la sécurité, la génération de nombres aléatoires (TRNG¹⁷ et PUF¹⁸) dans les circuits intégrés; utilisation du Machine Learning dans le domaine de l'intégrité des systèmes sur puce; test et fiabilité des circuits sécurisés, des technologies de mémoire émergentes.

- Les attaques par injection de fautes ciblant les circuits-intégrés sont déjà bien connues en particulier par l'exploitation du canal optique (par laser) et électromagnétique. De nouveaux moyens d'injection devraient permettre dans un avenir proche des attaques plus complexes, que ce soit des laser multi-spot (avec un nombre de spot supérieur à deux) et des rayons X. La compréhension fine des phénomènes physiques sous-jacents doit aussi être développée pour proposer des contre-mesures spécifiques notamment au niveau des nano/micro-technologies. Le développement de contre-mesures au niveau technologique (transistors résistants aux perturbations, écrans protecteur efficaces et peu coûteux, détecteurs ...) est donc un sujet d'importance avec encore peu d'équipes académiques engagées du fait de la complexité de maîtrise et d'accès aux dernières technologies microélectroniques.
- L'implémentation (logicielle et matérielle) efficace d'algorithmes de cryptographie est toujours un problème d'importance. Le développement d'outils de conception dédiés prenant en compte les aspects de sécurité est un point à renforcer.
- Les travaux concernant les TRNG et les PUF sont toujours d'actualité en particulier en ce qui concerne l'utilisation de nouvelles technologies. Sur ce point, un transfert important a eu lieu ces dernières années vers l'industrie, néanmoins de nombreux aspects théoriques et pratiques sont encore du domaine de la recherche académique. Ce point est d'autant plus important avec l'émergence d'approches matérielles "Open-Source" d'une part, et avec la fabrication de circuits intégrés se faisant actuellement hors Europe (en général en Asie) – il y a donc un impérieux besoin de lutter contre les chevaux de Troie (programme ou circuit matériel nuisible placé dans un programme ou circuit sain et effectuant des opérations malicieuses à l'insu de l'utilisateur).
- L'utilisation du Machine Learning dans le domaine de l'intégrité des systèmes sur puce est déjà une réalité industrielle, et son apport est considérable. On trouve par exemple des techniques d'apprentissage dédiées à la vérification de la conception des circuits intégrés, au diagnostic en volume réalisé après fabrication pour améliorer les rendements de production, ou à la mise en œuvre du standard ISO 26262 pour la sûreté fonctionnelle des systèmes embarqués dans les véhicules autonomes. Pour autant, les efforts doivent s'intensifier pour traiter d'autres problématiques de l'intégrité des systèmes dans lesquelles un grand nombre de données sont manipulées, comme par exemple la caractérisation de bibliothèques de cellules pour le test et le diagnostic de circuits numériques.
- Au fur et à mesure de l'évolution des architectures de circuits sécurisés, de nouvelles techniques de test pour ces circuits doivent être mises en place. Les conditions requises pour un test efficace (facilité d'accès aux nœuds internes d'un circuit) et pour une sécurité matérielle maximale (limitation des accès aux nœuds internes d'un circuit) étant à l'opposé, les challenges dans ce

¹⁷ True Random Number Generator

¹⁸ Physically Unclonable Function

domaine reste permanents et nécessitent l'apport régulier de solutions de test nouvelles et efficaces. Il en va de même pour les approches de la fiabilité.

- Les nouvelles technologies mémoires, comme les STT/SOT-MRAM, font appel à des processus de fabrication qui ne reposent plus exclusivement sur la technologie CMOS, mais qui combinent plusieurs technologies. Ces processus deviennent donc de plus en plus difficiles à maîtriser, et de nombreuses défaillances peuvent survenir lorsqu'il s'agit de produire des mémoires de grande capacité sur un substrat commun. Les stratégies de test doivent permettre d'appréhender ces défaillances de manière holistique. Par ailleurs, certains phénomènes associés à ces nouvelles technologies, comme par exemple la nature stochastique des commutations dans les mémoires magnétiques, soulèvent de nouveaux problèmes en termes de fiabilité qui, là encore, nécessitent l'apport de solutions nouvelles, efficace et faible coût.

Forces et faiblesses de la recherche scientifique française

Il existe une communauté académique importante dans cet axe au niveau national. Pour l'aspect cybersécurité, la sécurité *matérielle* (qui fait pleinement partie de cet axe) représente la seconde force en moyens humains déployés en cyber-sécurité au niveau des laboratoires CNRS après la cryptographie. La communauté en test et fiabilité est aussi très importante au niveau national. La force de ces acteurs académiques est d'être en lien fort avec des acteurs industriels de premier plan comme STMicroelectronics, Thales ou bien le CEA. De plus, les acteurs de la recherche ont la possibilité de diffuser les nouvelles connaissances dans de nombreuses formations supérieures et notamment les écoles d'ingénieurs françaises.

Les faiblesses dans cet axe, au-delà des faiblesses structurelles de la recherche publique française sont pour les plus importantes :

- le manque criant d'attractivité des acteurs académiques vis-à-vis des étudiants à bac+5 pour des poursuites en thèse de doctorat et en particulier pour les jeunes diplômés des écoles d'ingénieurs
- le manque d'équipes pluridisciplinaires alliant des compétences scientifiques complémentaires en électronique, informatique et domaines connexes comme les mathématiques appliquées par exemple. A cet égard, notons que la structuration de la recherche publique française divisée en sections (CNRS, CNU, ...) n'est pas en adéquation avec l'évolution de la complexité de nos sujets d'étude. Ce point est en partie couvert pour l'animation par les GdRs (dont SoC²), mais mériterait d'être étendu.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : INRIA-AOSTE, IRISA, INL, Laboratoire Hubert Curien, TIMA, Lab-STICC, LCIS, LIFL, LIP6, LIRMM, LORIA, LS2N, LTCI, ONERA, Telecom ParisTech, Verimag, XLIM

Autres : Ecole des Mines de Saint-Etienne

4. Objets connectés

Mots clés

Réseaux : Réseaux d'objets hétérogènes, Internet des objets, Communication sporadique,

Interaction objet/milieu : Systèmes Multi-physiques, Multi-échelle

Adaptabilité : Cognition, reconfigurabilité

Défis sociétaux : Privauté des données, Pervasivité et acceptation, Conception faible consommation, Autonomie, Impact sociétal, Écoconception

Domaines applicatifs : Territoire et Transports intelligents, Dispositifs et Implants biomédicaux, Usine 4.0

Les objets connectés (OC) sont aujourd'hui au cœur des enjeux économiques, sociaux et humains. Omniprésents dans nos sociétés numériques (au-delà des ordinateurs et autres *smart devices*), ils ont investi un large domaine applicatif permettant de répondre aux défis de demain en matière de mobilité, d'environnement, de santé et d'industrialisation. Leur diffusion, déjà importante, va s'intensifier et des estimations récentes prévoient plus de 100 milliards d'objets connectés pour la décennie à venir, soit plus de 15 fois la population mondiale.

On notera d'ailleurs que la notion d'IoT (Internet of Things) est en passe de devenir une notion d'IoE (Internet of Everything), où tout nouvel équipement, quel qu'il soit, aura des capacités de captation et de production de données. L'hyper-connectivité sera donc un enjeu important, et elle concerne tout à la fois des interactions *machine-to-machine*, *people-to-machine* et *people-to-people*.

Dans ce contexte d'évolution et de déploiement exponentiel, les défis qui sont à relever concernent plus particulièrement :

- Le passage à l'échelle des objets et méthodes
- La réduction de l'empreinte écologique du traitement, du transfert et du stockage des données qui deviendront massives
- La sécurité, au sens large, des données que ce soit dans leur traitement mais également pour leur transfert
- La diversité applicative des OC, chaque domaine apportant ses propres spécificités, besoins et contraintes en termes de latence, bande passante, criticité, densité, précision ...

Résoudre ces défis implique des questions relatives à la modélisation multi-physique / multi-échelle, à la conception des architectures (à haute efficacité énergétique) et des réseaux connectant les objets, aux mécanismes d'identification / adaptation et de compensation des environnements (complexes, sévères, etc.), et à la validation d'objets connectés soumis à de fortes contraintes liées aux domaines applicatifs et à leur environnement. Concernant la validation, des modèles de *jumeaux numériques* spécifiques à l'IoE devront être développés en vue de simuler et de vérifier le comportement des systèmes à base d'OC sur des scénarios applicatifs réels.

Par rapport aux axes identifiés dans le GdR SoC², l'axe Objets Connectés reste transverse, mais s'appuiera également sur les contributions concernant

- Les méthodes et outils pour la conception de systèmes en particulier pour cibler des architectures faible consommation
- L'intelligence artificielle qui va également diffuser au plus proche de la captation des données
- Des technologies émergentes qui devront soutenir l'augmentation des performances ainsi que la réduction de l'empreinte écologique de l'objet.

Thèmes actuels et évolutions attendues à l'horizon 2030

Aujourd'hui les objets connectés et l'IoT intègrent de plus en plus de mécanismes d'intelligence distribuée pour répondre à des défis sociétaux ou les problématiques d'intelligence collective sont prégnantes pour des tâches de détection, caractérisation et d'optimisation. Le traitement massif des données nécessite plus que jamais d'être au plus proche de la source d'émission pour réduire l'utilisation de la bande passante tout en garantissant la privauté. Le caractère intelligent de ces objets repose ainsi sur une dualité traitements intelligents et architectures embarquées adaptables avec des défis à différentes échelles. L'intégration sous contraintes de modèle d'IA se confronte à des problématiques de capacités de calcul qui sont forcément limitées dans un OC. Les challenges actuels et qui resteront probablement d'actualité concernant :

- La définition de modèles d'implémentation efficace du calcul d'inférence, incluant la réduction de la précision de calcul pour supporter celle-ci tout en conservant une qualité de convergence suffisante,
- La conception holistique (multi-échelle, multi-niveau) de réseaux d'OC adaptables à faibles empreintes énergétiques.

Les prochaines générations d'objets connectés s'orientent vers des dispositifs pourvus de **capacités d'apprentissage**. Des méthodes d'apprentissage pouvant être embarquées devront donc être développées et celles-ci devront avoir un impact limité sur la consommation énergétique de l'objet. Le déploiement, pour chaque application, de millions d'OC ayant à délivrer des services similaires, posera également la question de **l'apprentissage distribué**, pour lequel chaque OC pourra bénéficier de l'expérience des autres pour être capable de s'adapter à des situations nouvelles pour lui et ne correspondant pas aux situations auxquelles il a été confronté.

Dans ce contexte, la **sécurité de l'apprentissage** sera à adresser afin de s'assurer qu'un "mauvais" apprentissage d'un OC ne se propage pas dans le réseau d'OC, qui pourrait alors créer une vague de défaillances en chaîne au sein d'un réseau d'OC. Les aspects sécurité se posent également sur le transfert des données et donc aux interfaces des objets pour assurer la confidentialité des données et la privauté des utilisateurs au sein du réseau.

Bien entendu, l'aspect **consommation d'énergie de l'objet connecté** restera l'un des champs de recherche du domaine des OC, nécessitant la généralisation de nouveaux paradigmes de gestion d'OC comme le *normally-off computing* ou le *transient computing*. Les usages sont également un facteur important car les observations montrent qu'à chaque fois que l'on baisse la consommation d'énergie d'un OC, on permet l'implémentation de nouvelles applications, engendrant alors une nouvelle augmentation de la consommation au niveau de la toile des OC – vérifiant ainsi le [paradoxe de Jevons](#). Par ailleurs, dans plusieurs domaines où l'environnement est dit "*sévère*", ou "*critique*" (santé, défense, transport autonome, aéronautique, etc.) les OC intègrent des contraintes supplémentaires nécessitant de répondre à des défis spécifiques. Il s'agit par exemple pour le domaine de la santé d'un enjeu de prévention avec des systèmes implantés chroniquement, permettant d'identifier de manière précoce les pathologies et de les traiter au plus proche. Les OC sont ainsi confrontés à des problématiques d'adaptabilité système pour compenser les interactions avec le milieu.

Analyse des thématiques - défis importants et/ou émergents

Compromis haute performance versus empreinte écologique : le déploiement massif des objets pose implicitement la question de leur empreinte énergétique. Compte tenu de l'agilité de l'objet et des services associés, des mutations profondes au niveau de leur architecture seront nécessaires. Des techniques s'articulant autour d'un compromis entre approches distribuées et centralisées devront également être étudiées (données, placement des ressources de calcul et de contrôle. Ces évolutions devront permettre l'embarquement d'une plus grande capacité de calcul tout en garantissant un haut niveau de fiabilité, de protection et de résilience pour des applications numériques critiques (santé,

transport, industrie, etc.) Les thématiques de réduction de la consommation d'énergie, déjà abordées depuis de nombreuses années, resteront donc d'actualité pour la prochaine décennie. La définition d'architectures matérielles efficaces restera donc un pilier indispensable pour accompagner et soutenir l'augmentation des capacités de calcul, et l'accroissement des objets dans un monde hyper-connecté. Cela passera par la prise en compte d'un panel de techniques regroupant tout à la fois l'adaptation des équipements sur la base de leurs capacités de reconfiguration (logicielle et/ou matérielle), mais également l'exploitation de nouveaux paradigmes de calcul tel que, par exemple, le concept de *in-memory computing* qui permet de réduire le mouvement des données très consommateur d'énergie au sein des architectures. Enfin, en comparaison des implémentations logicielles sur CPU ou GPU classiques, des implémentations matérielles n'ont plus à démontrer leur efficacité, notamment au travers des circuits reconfigurables de type FPGA. Toutefois, leur diffusion à grande échelle passera par une simplification (déjà en cours) de leur exploitation au niveau applicatif.

- **Passage à l'échelle des objets et des données** : Compte tenu de la quantité de données qui est d'ores et déjà produite par les objets connectés, la pertinence des informations captées et remontées vers les éléments de décision (très souvent situés dans le cloud) devient de plus en plus critique. Au fur et à mesure du déploiement des millions/milliards d'objets, la pression sur l'analyse *in situ*, par l'objet lui-même, s'accroît en vue d'éviter d'encombrer à la fois le réseau de communications et le cloud par un flux de données brut inexploitable. Ainsi, une coopération efficace entre le *edge computing* (*near-sensor computing*) et le *cloud* pour une meilleure exploitation des données et ressources de calcul et de communication est nécessaire. Il est primordial, dès à présent, d'imaginer de nouveaux paradigmes de traitement afin d'augmenter la pertinence des données recueillies tout en répondant à des contraintes de temps sévères, de consommation énergétique, de prise de décision, de conception sur le plan environnemental (écoconception, biodégradabilité, ...) et d'acceptabilité. L'embarquement de capacité d'IA au sein des objets, en conjonction avec d'autres techniques comme le calcul distribué et le *in-network computing*¹⁹, ou encore le calcul approximé et la quantification, sera donc l'une des solutions incontournables pour assurer un passage à l'échelle.
- **Réseaux d'objets** : Si le traitement local des données pour en extraire une information pertinente permettra de réduire la bande passante utile pour chaque objet, cette réduction ne suffira pas pour réduire le trafic sur les réseaux qui devront assurer la connectivité d'un plus grand nombre d'objets. La convergence des réseaux, en partie couverte par la 5G, pourrait être la solution, mais les besoins étant extrêmement hétérogènes, des problématiques liées à la définition de nouvelles technologies de communication mobile (6G, THz ...) et de nouveaux protocoles réseaux (BLE²⁰, LPWAN²¹ ...) devront être poursuivies dans les prochaines années.
- **Sécurité des objets/ données** : La multiplication des objets s'appuie en partie sur leur faible coût, qui est parfois obtenu en embarquant un faible niveau de sécurité afin de limiter les besoins en termes de capacité de calcul. Toutefois, dans un contexte IoE, chaque objet peut devenir une

¹⁹ In-Network Computing est un paradigme de calcul émergent réalisant des opérations de calcul au sein du réseau (par exemple le réseau *edge-cloud*). Il s'appuie sur les dispositifs (tels que des FPGA ou des ASIC programmables) qui existent déjà dans le réseau et qui sont déjà utilisés pour acheminer le trafic, minimisant ainsi la surcharge du calcul et en déchargeant les calculateurs dédiés aux extrémités du réseau. De plus, cette approche réduit la charge du trafic dans le réseau, car les transformations opérandes-résultats réduisent le volume des données et permettent de mettre fin aux transactions lorsque les données traversent le réseau.

²⁰ Bluetooth Low Energy

²¹ Low-Power Wide Area Network

porte d'entrée vers le système d'information dans son ensemble. L'hyper-connectivité de ces objets ne pourra alors être assurée que si des mécanismes de certification de confiance sont délivrés et ceci de façon dynamique pour contraindre des pénétrations dans les systèmes. L'identification des intrusions sera de plus en plus nécessaire pour se prémunir d'attaques malveillantes. La sécurité sera alors en enjeu majeur à tous les niveaux, allant des aspects matériels de l'objet jusqu'aux couches logicielles hautes incluant les échanges réseaux.

- **Digital Twin IoT** : Le *jumeau numérique* représente un outil de modélisation et de simulation à l'identique d'un système donné. Cette simulation est basée principalement sur des données réelles. Pour certains cas d'études, tel que la régulation du trafic routier pour une ville intelligente ou encore la détection anticipée des dangers (accidents, crimes, bouchons, stationnement, etc), des outils de création et de gestion de *jumeaux numériques* seront essentiels afin de pouvoir créer différents scénarios permettant de mettre en place une ou plusieurs solutions efficaces. Ces techniques seront à développer pour étudier les comportements (dans toutes ses dimensions) d'un réseau de capteur, en vue par exemple de vérifier son comportement dans les cas d'usages classiques mais également dans des cas d'usages correspondant à des modes dégradés et/ou des attaques malveillantes.
- **Privauté, confiance, acceptabilité et appropriation** : L'évolution vers l'IoE s'accompagne d'une hyper-connectivité dont l'utilisateur doit comprendre l'intérêt. S'il paraît relativement logique que ces objets sont acceptés lorsqu'ils rendent un service recherché par un utilisateur, il existe un risque non négligeable de rejet si ces objets s'insinuent dans le quotidien de l'utilisateur pour le tracer ou le modéliser, à son insu, et avec un objectif échappant totalement à son contrôle. Pour accorder sa confiance à un objet, l'utilisateur pourra exiger une connaissance précise et compréhensible du niveau d'intrusion de celui-ci dans son environnement personnel et professionnel. L'acceptabilité et l'appropriation de l'objet passera donc par une éthique sociétale dans le développement et le déploiement de celui-ci.
- **Domaines impactés par l'IoE** : L'ensemble des points soulevés ci-dessus n'est pas spécifique à un domaine d'application, mais pourra s'appliquer avec des pondérations différentes. Il est par exemple évident que dans un contexte médical, la sécurité des données captées puis transmises sera primordiale pour que le couple patient-médecin soit en confiance et accepte l'intrusion, in-vivo, d'objets pour surveiller des paramètres physiologiques et/ou activer la libération de molécules dans le cadre d'un traitement à distance. Dans un tout autre domaine qu'est la sécurité routière, la problématique de la consommation énergétique sera sans aucun doute moins forte pour des équipements à destination de véhicules disposant d'une source d'énergie assez confortable. Le domaine du transport intelligent, impliquant de l'interaction entre véhicules ou entre véhicule et infrastructure, les problématiques réseaux et d'ubiquité seront prédominantes pour assurer la fluidité du trafic et surtout la sécurité de celui-ci. Les réseaux d'IoE seront aussi déterminants pour la mise en place de *jumeau numérique* pour des domaines très variés tels que l'usine du futur, la ville intelligente, le transport intelligent, etc. Pour être exploitable et permettre des études précises, le *jumeau numérique* s'appuie sur des données qui sont captées dans un environnement réel. Les flots de données remontant des objets seront donc tout à fait pertinents pour ce domaine applicatif en plein essor.

Forces et faiblesses de la recherche scientifique française

Un grand nombre de laboratoires français dispose de compétences couvrant tout ou partie des thématiques relevant des objets connectés, mais il n'y a pas, au jour d'aujourd'hui, de pôle d'excellence visible en France.

Pour parvenir à adresser l'ensemble des défis identifiés ci-dessus, la complémentarité des laboratoires et des équipes de recherche est indispensable. Depuis de nombreuses années, des collaborations nombreuses existent entre les chercheurs des sciences "dures" pour adresser ces challenges, mais le déploiement massif devra prendre en compte des aspects liés aux sciences sociales et humaines. Il sera donc nécessaire d'inclure des collaborations mêlant sciences "dures" et sciences humaines et cela impactera sans aucun doute les objectifs classiques que nous adressons au niveau du GdR SoC².

Par ailleurs, les domaines applicatifs seront sans aucun doute des catalyseurs de compétences pour répondre au mieux aux spécificités des usagers. En effet, les domaines du transport intelligent et de la santé, pour ne prendre que ces deux exemples, ne privilégieront pas les mêmes objectifs mais s'appuieront sur des contributions très large couvrant les défis cités ci-dessus.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : C2N, CITI, ETIS, FEMTO-ST, IM2NP, ICUBE, IETR, IJL, IMS, INL, IRISA, Lab-STICC, LAMIH, LEAT, LIP6, LIRIS, LIRMM, TIMA.

5. Technologies du futur

Mots clés

Nouveaux paradigmes de calcul, de mémorisation et de communication

Hybridation technologique, intégration 3D, système hétérogène

Quantique

Calcul neuro-inspiré

Électronique durable

Au cours des dernières décennies, les systèmes électroniques ont envahi notre vie quotidienne avec l'avènement des applications connectées. Cette évolution s'est appuyée sur la mise en place d'infrastructures de type objets connectés, avec des quantités massives de données transférées et traitées dans le Cloud sur des systèmes de calcul haute performance. Cette tendance s'accélère, avec une prévision de plusieurs milliards d'appareils connectés d'ici 2030, tout en soulevant des questions de souveraineté et de sécurité. Le corollaire de cette évolution, est l'augmentation de la consommation énergétique globale résultant de la circulation des données, de leur stockage et de leur traitement massif dans un environnement sécurisé. Cette évolution doit également prendre en compte l'urgence écologique et doit maintenant s'appuyer sur des solutions durables avec une faible empreinte carbone et un contrôle éthique et sécurisé de l'utilisation des données personnelles.

Dans ce contexte, l'intégration de technologies émergentes dans la conception des systèmes électroniques doit évoluer en embarquant l'intelligence au plus proche des capteurs, avec la mise en œuvre :

- de nouvelles méthodologies de conception (axe Méthodologies et outils)
- de nouveaux outils en matière de co-simulations multi-physiques (axe Frontières et interfaces cyberphysiques)
- de nouveaux paradigmes de calcul (axe Calcul embarqué haute performance/ axe IA embarquée)
- de nouvelles solutions de sécurité des informations et des systèmes (axe Sécurité et intégrité des systèmes embarqués)

Thèmes actuels et évolutions attendues à l'horizon 2030

Les thèmes pivots de l'axe technologies du futur peuvent être regroupés autour des trois thèmes suivants :

- **Les futurs systèmes de communication sans fil** : ils devront permettre d'augmenter les débits tout en réduisant la consommation d'énergie lors des communications. Ces systèmes se positionnent au-delà de l'évolution actuelle prévisible "6G", et feront appel à des systèmes hybrides électroniques/optiques, visant les communications optiques (THz) en espace libre.
- **Les futurs cœurs des objets connectés** : ils devront intégrer une électronique de plus en plus hétérogène pour permettre une interaction accrue avec le monde physique. L'intégration monolithique de capteurs faisant appel à différentes physiques est attendue au-delà des systèmes actuels de types MEMS²², avec des interfaces vers l'optique, le fluide, etc... Le traitement des données, tout ou en partie dans l'objet est aussi attendu, avec le déploiement de moteurs d'IA pour un traitement complet ("*near-sensor computing*", "*edge computing*") ou partiel ("*in-network computing*") dans l'objet.
- **Les futurs systèmes de calcul haute performance** : ils devront permettre une efficacité énergétique optimale, en autorisant un traitement massivement parallèle de l'information. De

²² Micro-Electro-Mechanical Systems

nouveaux paradigmes de calcul, centrés sur les données devront être mis en avant ("*non Von-Neumann computing*"). Dans le même temps, l'essence même des calculateurs devra évoluer avec l'émergence du calcul quantique et de son interfaçage avec des systèmes conventionnels.

Analyse des thématiques - défis importants et/ou émergents

Le premier obstacle identifié concerne les futurs systèmes de communication sans fil, qui doivent permettre d'augmenter les débits tout en réduisant la consommation d'énergie lors des communications. Cet obstacle est lié à la convergence de l'électronique et de l'optique pour permettre la propagation en espace libre des communications optiques. Ces technologies, aux frontières de l'électronique et de l'optique, sont les technologies de communication THz. La connaissance approfondie de leurs caractéristiques (puissance d'émission et de réception, forme d'onde, format de modulation, etc.) permettra d'extrapoler les performances des futurs réseaux locaux. Au-delà, un autre aspect concerne la mise en place de communications optiques intra-puce dans un même système afin de réduire le coût énergétique des communications.

Le deuxième goulot d'étranglement identifié est la co-intégration de technologies hétérogènes (capteurs, optique, fluide, etc.), pour augmenter la fonctionnalité des technologies CMOS actuelles (FDSOI²³, FinFET) ou à venir (GAA-FET²⁴ latéral ou vertical), et permettre ainsi une meilleure efficacité énergétique de l'interaction cyber-physique. Pour permettre cette conception hétérogène, et pour appréhender les compatibilités technologiques à différentes échelles, de nouveaux outils de simulation multi-physique portant sur les niveaux des procédés de fabrication, des composants, des circuits et des systèmes seront nécessaires. La prise en compte de l'intégration 3D monolithique de technologies hétérogènes ou de l'empilement hétérogène des puces, ou encore de l'intégration 2.5D au niveau de la mise en boîtier, est l'une des principales pistes de développement de ces systèmes augmentés.

Le troisième obstacle identifié est lié aux architectures haute performance, centrées sur le calcul ("*Von-Neumann computing*"), qui atteignent leurs limites pour les applications nécessitant un traitement massif des données (par exemple réseaux neuronaux artificiels). Cette architecture révèle clairement un double problème : la latence dans les transferts de données entre des technologies hétérogènes (cœurs de calcul, mémoires volatiles et non volatiles) et la consommation d'énergie associée au mouvement des données. Trois voies d'exploration à envisager, qui ne s'excluent pas mutuellement, peuvent être envisagées, pour permettre le choix des meilleures technologies :

- Le renversement de l'architecture, avec des systèmes centrés sur les données, et des approches de calcul proche- ou en-mémoire. Les explorations liant les couches technologiques jusqu'aux couches logicielles devront être menées pour favoriser les choix technologiques.
- La mise en œuvre de nouveaux paradigmes de calcul (stochastique, approximatif, hyperdimensionnel, arithmétique non-conventionnelle etc.) tirant le meilleur parti de nouvelles technologies (quantique, mémoire, etc...) pour réduire l'empreinte des données et augmenter la robustesse du système.
- La conception d'architectures neuro-inspirées, mettant en œuvre, selon les applications, un calcul adiabatique avec une fréquence de fonctionnement assez lente, proche de celle du cerveau humain. Ces architectures pourraient être déployées en périphérie pour limiter le transfert de données critiques en favorisant le traitement local ou dans le réseau.

²³ Fully-Depleted Silicon on Insulator

²⁴ Gate-All-Around Field Effect Transistor

Forces et faiblesses de la recherche scientifique française

L'idée principale de tous ces développements est de promouvoir l'interaction cyber-physique, avec un traitement local et résilient des informations, centrées sur les données. Cette approche permettra de réduire au minimum la quantité de données à échanger grâce à des systèmes de communication plus efficaces sur le plan énergétique et de partager le traitement de l'information entre objet, réseau et cloud. Les architectures de demain ne seront plus exclusivement CMOS et numériques, mais feront appel à de l'hybridation ou de l'hétérogénéité technologique, une des voies clairement identifiées est le passage massif en 3D des circuits avec l'utilisation de la technologie la plus adaptée à chaque fonction (calcul, mémorisation, communication) ; ces fonctions ayant tendance par ailleurs à fusionner (comme par exemple avec l'émergence de calcul analogique en mémoire) et à faire appel à de vraies ruptures technologiques comme le quantique.

Le succès de la recherche en France dans ces domaines passe :

- D'une part par la maîtrise et la capacité à fabriquer, caractériser et modéliser les nano composants permettant d'élaborer ces nouvelles architectures système, grâce au Plan Nano 2022 et au réseau national de grandes centrales de technologie (Renatech)
- D'autre part à déployer ces nano-composants à l'échelle de circuit au travers de réalisations matérielles permettant de calibrer des modèles abstraits de ces circuits pour passer à l'échelle d'architectures complètes

Les laboratoires français impliqués dans ces axes de recherche couvrent la quasi-totalité des compétences scientifiques nécessaires au développement de cette chaîne de valeur, à l'exception de l'élément clé permettant le passage à l'échelle du nano-composant unitaire au circuit complexe. La maîtrise de cet élément clé passe pour l'instant par la mise en place de collaborations ponctuelles, à l'échelle nationale avec le CEA-Léti et à l'échelle européenne avec l'IMEC. En effet, seules ces deux centrales technologiques permettent, en Europe, le passage à l'échelle du nano-composant au circuit complexe en agrégeant des technologies hétérogènes en 3D, à l'instar des centres de recherche multi-laboratoires aux USA ou interuniversitaires en Chine. Pour accroître le succès et la visibilité des nouveaux nano-composants développés dans les laboratoires du CNRS d'une part et d'autre part favoriser la maîtrise des technologies de demain et leur transfert dans l'industrie, la mise en place d'un partenariat institutionnel pérenne entre le CNRS et ces centrales technologiques nous paraît indispensable.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : C2N, IEMN, IES, IETR, IM2NP, IMS, INL, LAAS, LIRMM, SPINTEC, TIMA

6. Méthodes et outils

Mots clés

Conception Assistée par Ordinateur, Electronic Design Automation, Electronic System Level

La complexité des systèmes électroniques atteint des sommets avec des centaines de cœurs de calcul et des dizaines de milliards de transistors sur une seule puce. La performance des systèmes ne vient plus simplement d'un plus grand nombre de ressources mais d'une meilleure utilisation de celles-ci, de la spécialisation des composants entraînant encore une plus grande hétérogénéité, ou encore de nouveaux paradigmes de calculs comme le calcul en mémoire ou le calcul approximatif. Ces techniques nécessitent des méthodes et des outils pour leur mise en œuvre. L'augmentation du niveau d'abstraction et l'unification de la représentation des composants logiciels et matériels du système permettront de concevoir et programmer plus facilement des systèmes numériques. Le développement de l'intelligence artificielle offre de nouvelles possibilités dans la modélisation et l'optimisation des systèmes complexes qui sont à intégrer dans les outils et méthodes de développement. Enfin, la maîtrise de la consommation d'énergie des circuits est un enjeu majeur des prochaines années. Les méthodes et outils permettant la gestion de la consommation d'un système dans sa globalité sont encore à perfectionner.

Ici, les défis concernent :

- Les méthodes de conception pour les architectures émergentes comme le calcul approximé, le "*in-memory computing*" et l'intelligence artificielle
- L'intégration de l'intelligence artificielle pour l'optimisation de circuits et la modélisation d'architectures complexes
- L'augmentation du niveau d'abstraction pour les architectures hétérogènes et
- La modélisation et l'optimisation de la consommation d'énergie d'un point de vue architecture et système.

Thèmes actuels et évolutions attendues à l'horizon 2030

La fin de la loi de Moore ne signifie pas encore la fin de la technologie CMOS mais d'autres technologies encore à l'étude prendront la relève (cf. axe 5 Technologies du futur). Sans être entièrement remises en cause, les méthodes actuelles doivent être revisitées pour intégrer des nouvelles contraintes (fiabilité, sécurité, consommation). En outre, l'intégration de plusieurs composants unitaires pour réaliser un système complet est souvent laissée à l'expertise de l'ingénieur et réalisée manuellement. L'automatisation de l'intégration et de l'interfaçage des composants, avec les tests et vérifications automatiques deviennent un réel besoin.

La montée en puissance des architectures hétérogènes et l'augmentation de leur taille et complexité nécessite un passage à l'échelle des méthodologies de conception vers des systèmes complexes. Cela nécessite l'abstraction de ces architectures par des techniques de virtualisation, de synthèse de haut-niveau et d'intégration des systèmes hétérogènes. Cette augmentation en complexité représente un coût en ingénierie important, et la recherche doit nourrir autant que s'appuyer sur les technologies open source logicielles et matérielles pour explorer ces systèmes complexes.

Le développement de l'intelligence artificielle pose de nombreux défis quant à son intégration sur des architectures spécialisées, mais offre aussi de nouvelles opportunités pour la conception d'outils de haut niveau. L'intelligence artificielle permet la modélisation et l'optimisation de systèmes complexes, en complément des méthodes de recherche opérationnelle utilisées dans les outils actuels. Elle devrait ouvrir de nouvelles opportunités que ce soit dans la conception de circuits à bas

niveau, la conception d'architecture spécialisées de manière automatique, la modélisation et l'optimisation d'applications à haut niveau.

La performance énergétique est devenue le facteur clef des systèmes numériques. Les contraintes énergétiques déterminent les architectures de calcul haute performance où la consommation est strictement limitée, pour des raisons économiques, à quelques mW. Les objets connectés doivent, quant à eux, garantir un fonctionnement sur accumulateur durant des années voire des décennies. Dans ce contexte, un gain d'un facteur 10 en énergie par opération est souhaitable au niveau système dans la prochaine décennie pour l'ensemble des systèmes numériques. L'efficacité énergétique peut devenir une métrique pour les méthodes, avec des approches centrées sur l'énergie, ou sur la mémoire.

Les outils de simulation sont et resteront des acteurs incontournables. Le traditionnel compromis entre temps de simulation et précision de la simulation est exacerbé par la complexité des systèmes à concevoir. L'approche basée sur les traces d'exécution atteint ses limites lorsque l'ajout d'un composant impacte lui-même les traces. Une approche holistique au niveau système est souvent la seule approche possible, mais la précision au niveau cycle d'un système complet, incluant plusieurs dizaines de processeurs et un système d'exploitation par exemple, est tout simplement hors d'atteinte. La perte en précision de mesure pose alors la question de la pertinence des résultats lorsqu'il est question d'évaluer l'apport (temps d'exécution, consommation énergétique) d'un composant dans le système global. Les solutions basées sur des accélérateurs matériels, émulateurs, ou "*hardware in the loop*", offrent des perspectives séduisantes si ce n'est les seules crédibles. D'autres approches par *model-checking statistique* peuvent être envisagées mais nécessitent encore des développements pour la prise en compte d'architectures non-composables.

Analyse des thématiques - défis importants et/ou émergents

Parmi les thématiques en émergence, quatre familles de méthodes de conception sont à suivre :

- Pour l'IA embarqué
- Pour la cybersécurité
- Pour les approches approximées ou proche mémoire
- Pour le matériel open-source

Concernant les méthodes pour l'IA embarquée, de nombreuses pistes restent encore à étudier, incluant les systèmes d'IA peu ou non supervisés, à apprentissage continu, ou légers "par nature".

Concernant les méthodes pour la cyber-sécurité, certains mécanismes de protection contre des attaques connues comme les canaux cachés pourraient être directement intégrés et proposés au concepteur sous forme de génération automatique de composants.

Concernant les méthodes pour les nouvelles approches de calcul comme le calcul approximé, celles-ci peuvent se placer au niveau algorithmique, ou au niveau architectural. La "*transprecision*" permet par exemple d'approximer les nombres flottants en proposant un rapport précision/gain énergétique très intéressant. Une autre approche prometteuse englobe le calcul "proche-mémoire" ou "dans" la mémoire. Ces deux possibilités cassent la couche logicielle traditionnelle et il convient de définir de nouvelles méthodes, voire même de nouveaux langages de programmation, pour exploiter pleinement cette piste très prometteuse. L'interaction calcul-mémoire fait l'objet de nombreuses recherches notamment à l'international. L'arrivée concomitante des technologies mémoires non-volatiles rend ces recherches d'autant plus disruptives.

Concernant les méthodes pour le matériel open-source, il convient de suivre de près les nombreuses initiatives qui émergent dans de nombreux pays, y compris au niveau européen. Le jeu d'instructions ouvert RISC-V est l'exemple criant d'un engouement sans précédent pour l'open source matériel. Ce mouvement peut (et doit pour garder une forme de complétude) être accompagné par des logiciels

open source permettant de concevoir un circuit. D'un point de vue sécurité, la connaissance et la couverture de la totalité de la chaîne de conception d'une puce permet de se prémunir de toute insertion de matériel non désiré. Pour les académiques, le matériel open-source permet d'avoir accès aux mêmes ressources que les industriels sans les prix parfois rédhibitoires de certains outils, ou les clauses ne permettant pas de publier les résultats de recherche. Pour les industriels, le matériel open-source permet une certaine indépendance vis-à-vis de certains fournisseurs, de bénéficier d'une factorisation des efforts, de garder une maîtrise, et d'assurer la pérennité des équipements ou produits. À l'instar d'une suite logicielle comme Linux qui n'empêche en rien une activité économique, le logiciel pour le matériel open source pourrait se tourner vers un nouveau modèle économique.

Il existe de nombreuses thématiques existantes qu'il convient de maintenir ou de renforcer. Commençons par citer les approches pour les circuits reconfigurables (FPGA, CGRA²⁵), embarqué ou dans le cloud, ou pour les GPUs, sans oublier les circuits dédiés, et en particulier pour l'IA. Les méthodes pour les systèmes hétérogènes restent plus que jamais d'actualité, notamment celles considérant les signaux mixtes, pour les technologies ou les communications émergentes dans la puce comme l'optique ou la radio. Le domaine encore balbutiant, mais bien existant, spécifique à l'apprentissage profond doit être renforcé par des outils d'optimisation et d'automatisation de réglages de paramètres, d'accélération de l'inférence et de l'apprentissage. Un autre domaine à fortifier est celui de la vérification, incluant les tests (génération de vecteurs de tests, BIST²⁶) et la vérification formelle (preuve). Enfin, de nombreuses autres approches sur des problématiques bien ancrées restent à étudier : temps-réel, basse consommation, ou encore compilation au sens large.

Les méthodes traditionnelles de la recherche opérationnelle (RO) (heuristiques, méta-heuristiques, méthodes exactes) souffrent de la concurrence des méthodes basées IA/ML. La communauté est poussée à tout résoudre avec l'IA/ML, mais ces méthodes ne conduisent pas systématiquement à bien identifier et formaliser les problèmes. Il convient de rester attentif le temps de disposer de suffisamment de données concernant le coût en temps et en énergie des méthodes IA/ML, et d'adapter la bonne méthode en fonction du problème. Il est surtout primordial de ne pas perdre les connaissances et compétences acquises en RO traditionnelle afin de pouvoir les réutiliser à l'horizon 2030 lorsque les méthodes basées IA/ML auront été intensivement étudiées par la communauté.

Forces et faiblesses de la recherche scientifique française

La grande diversité des laboratoires travaillant dans les méthodes et outils de conception offre un environnement fertile au développement de ces travaux. Cette diversité est d'une part dans les domaines abordés, depuis la conception de circuits jusqu'à la programmation d'architectures hétérogènes complexes, et d'autre part dans les méthodologies utilisées en collaboration avec d'autres groupes de recherche comme la recherche opérationnelle, ou le développement fort de l'intelligence artificielle dans l'ensemble de la communauté française.

La grande faiblesse dans ce domaine est le manque de souveraineté sur une grande partie des outils développés, avec l'utilisation d'outils de conception et d'architectures fermées où les contributions à la méthodologie sont limitées. La simple utilisation de ces technologies de pointe (sans parler de quelconque contribution scientifique) n'est possible qu'à condition de respecter des clauses strictes, nécessitant parfois l'aménagement de postes de travail ad-hoc, déconnectés du réseau d'entreprise, dans une zone d'accès réglementée, moyennant un coût parfois restrictif malgré les programmes spécifiques mis en place. Les acteurs académiques français sont majoritairement déconnectés des

²⁵ Coarse-Grain Reconfigurable Architecture

²⁶ Built-In Self-Test

méthodes et outils de synthèse de bas-niveau, les grands pourvoyeurs étant implantés dans d'autres pays. Le manque d'accès aux outils de synthèse bas niveau, sous-entendu pour des architectures matérielles seulement, n'est pas rédhibitoire. La communauté française s'oriente vers des outils dits de "haut niveau", couvrant un spectre plus large incluant du matériel et du logiciel.

Le développement d'un mouvement pour le matériel et le logiciel ouvert pour l'architecture en Europe constitue une réelle opportunité à saisir, et une implication forte de la recherche française permettrait à la fois de renforcer les ressources pour le développement de la recherche en méthodologie et outils de conception et regagner la souveraineté dans ce domaine. L'accès au cœur même des outils, à travers l'accès au code source, devient une condition sine qua non pour pouvoir contribuer scientifiquement. Mais la complexité des techniques actuelles est telle qu'une initiative individuelle représenterait une charge de travail difficilement soutenable et presque impossible à assumer pour une seule entité. Il faut nécessairement passer par une mise en commun et un développement collectif pour espérer rivaliser avec les solutions existantes et bien ancrées héritant de plusieurs décennies d'un travail acharné et de compétences rares.

Enfin, la complexité des outils demande une montée en compétence qui nécessite du temps. Au mieux, l'enchaînement des contrats courts essouffle une équipe qui doit continuellement renouveler ses forces vives. Au pire, l'équipe doit abandonner tout développement par manque de moyens.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

UMR-CNRS : CITI, CRISAL, ENSTA, ParisTech, IETR, INL, INRIA Paris, Institut Pascal, IRISA/INRIA Rennes, IRIT, Lab-STICC, LAMIH, LIAS, LIGM, LIP, LIP6, LIRMM, LS2N, LTCL, TIMA, Verimag

7. IA et systèmes embarqués

Mots clés

Techniques : IA distribuée, coopération cloud-edge, matériel neuro-inspiré, représentation des données

Défis : efficacité énergétique, sécurité des données

Applications : conception intelligente des systèmes, interfaces IA-humain

L'Intelligence Artificielle est une des sciences qui marquera le plus profondément ce siècle, l'humanité progressant à pas de géants dans ce domaine. Les progrès incessants en apprentissage statistique par exemple catalysent et ouvrent des perspectives nouvelles dans tous les champs disciplinaires comme les sciences humaines et sociales, la biologie, l'environnement, la médecine etc. L'avènement de l'IA telle qu'on la connaît aujourd'hui mais également les progrès ininterrompus en apprentissage sont intimement liés à la disponibilité d'architectures de calcul toujours plus performantes. Travailler sur des technologies de rupture permettant l'augmentation sensible de l'efficacité énergétique de l'implantation de tels algorithmes constitue un défi de premier plan non seulement pour des applications embarquées mais également dans le but de réduire l'empreinte carbone préoccupante de ces technologies. Au-delà des aspects architectures matérielles et logicielles, l'ensemble des thèmes du GdR SoC² témoignent des bouleversements profonds induits par l'IA, au niveau par exemple des méthodologies exploitant l'IA, des systèmes cyber physiques, de la sécurité, des technologies émergentes pour le neuromorphique etc.

Thèmes actuels et évolutions attendues à l'horizon 2030

L'IA est omniprésente dans l'intégralité des thèmes couverts par le GdR SoC², qu'elle soit une composante de la vocation applicative, utilisée en amont en phase de conception ou vue comme une préoccupation au sens de propriétés de circuits telles que la certifiabilité ou la sécurité. On peut au sens des travaux de recherche menés distinguer les grandes directions suivantes :

IA, Architectures et Applications : de nombreuses recherches sont focalisées sur la dimension applicative et en particulier pour le "Edge AI", qui prône une décentralisation de l'IA, en abordant divers aspects ; logiciels, matériels et à l'interface tels que :

- Algorithmes, adéquation algorithme / architecture et architectures dédiées pour certaines applications contraintes, y compris techniques d'optimisation de portage de réseaux de neurones
- Conception d'accélérateurs dédiés à certains types de réseaux, les réseaux convolutionnels étant particulièrement populaires pour les applications de vision, mais aussi d'autres types de réseau récurrent (RNN²⁷, LSTM²⁸ etc.)
- Apprentissage embarqué / incrémental pour autoriser une spécialisation automatique dans le cas d'applications à coloration IoT / Edge AI

Composants et Technologie : l'efficacité énergétique étant une préoccupation prégnante, nombre de travaux visent à repenser l'approche conventionnelle largement basée sur l'utilisation de matériel numérique émulant des modèles de neurones "conventionnels" via une approche essentiellement orientée algèbre linéaire. On peut citer à ce titre un engouement particulier pour les réseaux à spikes (SNN²⁹) à la frontière avec les neurosciences mais aussi d'autres travaux inspirés de la physique fondamentale tels que les *Energy Based Models* (EBM). Si certains travaux utilisent du matériel

²⁷ Recurrent Neural Network

²⁸ Long Short-Term Memory

²⁹ Spiking Neural Network

numérique ou analogique, nombre d'approches se positionnent en rupture en utilisant des technologies émergentes, avec des perspectives d'amélioration en efficacité énergétique très significatives.

IA, Sécurité, Fiabilité, Test : dans le domaine de la sécurité l'IA est un sujet d'actualité à différents titres, de par les attaques connues sur des IA état de l'art en vision (*adversarial attacks*) mais aussi de par la puissance des modèles qui obligent à repenser les contre-mesures usuelles contre les attaques logicielles et matérielles. Dans le domaine du test également l'IA occupe une place de premier plan comme outil d'amélioration d'analyse, de conception jusqu'aux importantes considérations sous-jacentes à la fiabilité des systèmes, telles que certifiabilité et explicabilité.

IA pour la CAO : l'IA, avec les approches ML incluant du raisonnement ainsi que des approches plus cognitives comme les réseaux Bayésien, est également considérée sur le volet méthodologique, en phase de conception, pour divers aspects tels que l'estimation des performances, le diagnostic en extension des modèles statistiques conventionnels "historiques" tels qu'analyse de Monte-Carlo etc. Elle est également considérée comme un outil possible pour l'amélioration des performances d'algorithmes à différents niveaux (placement, routage) mais également pour l'exploration d'espaces de conception par le biais d'utilisation de technique état de l'art en IA (IA génératives par exemples).

Analyse des thématiques - défis importants et/ou émergents

Les challenges majeurs de l'IA embarquée et de l'IA pour l'embarqué actuellement affrontés peuvent être décomposés comme suit :

Souveraineté, sécurité, safety, confidentialité, défis éthiques : les défis dans ce domaine sont nombreux et dépassent le seul cadre du GdR. L'utilisation massive de l'IA sur des données confidentielles crée notamment de nouvelles surfaces d'attaques sur ces données en termes de confidentialité, intégrité et disponibilité. Cela se traduit en IA embarquée par de nouvelles vulnérabilités sur lesquelles développer des contre-mesures notamment en termes d'architecture de calcul et de substrats sera nécessaire. L'IA distribuée proche des capteurs peut en revanche fournir des solutions à certains problèmes de confidentialité, les données brutes des capteurs n'étant plus transférées sur le réseau. De façon générale, le support matériel de calcul peut intervenir sur ces sujets en tant que frein ou accélérateur, rendant indispensables des recherches interdisciplinaires sur ces domaines.

Efficacité énergétique, développement durable : L'efficacité énergétique a, dans un cadre embarqué, deux objectifs majeurs :

- Rendre possible en embarqué des tâches jusqu'ici cantonnées aux systèmes conventionnels
- Réduire les besoins en termes de production et de stockage d'énergie.

Le développement durable présente, outre la consommation d'énergie à l'utilisation de l'IA, deux autres aspects qui s'appliquent à l'IA comme aux autres formes de calcul : l'utilisation de matières premières et d'énergie à la fabrication du système, et le coût environnemental du recyclage en fin de vie du système. Les recherches en ce domaine au sein du GdR SoC² sont très vastes, allant notamment de l'optimisation des hyper-paramètres des réseaux à la représentation des données, au calcul approximé, et aux nouveaux substrats de calcul.

IA "extrême" (*Beyond CMOS*) : l'efficacité énergétique de l'IA est aujourd'hui poussée, sur les algorithmes les plus en pointe, au maximum des capacités du substrat CMOS et des architectures matérielles. Un effort sans précédent ayant été apporté en ce sens au cours des dix dernières années, l'IA permet d'évaluer les capacités maximales atteignables par les substrats actuels. Malgré ces efforts d'optimisation, les capacités de modélisation de l'IA restent aujourd'hui limitées par les puissances de calcul disponibles. L'IA est donc un formidable vecteur d'amélioration pour les technologies post-CMOS telles que le calcul quantique, ou le calcul neuromorphique.

Interface IA/humain "augmenté" (capteurs) : la capacité de l'IA embarquée à exploiter des données brutes de capteurs pousse également à créer de nouveaux capteurs intelligents intégrant l'extraction de données et leur analyse. Au-delà des aspects capteurs, l'IA embarquée devra s'adapter à moyen terme aux évolutions probables de l'IA en termes de collaboration humain-machine, assistant les personnes dans des tâches de plus en plus complexes.

IA assistance (CAO) : en ce sens, l'intégration de l'IA dans la conception des systèmes embarqués elle-même ouvre de nouvelles opportunités en termes d'optimisation des circuits mais également d'adaptabilité des systèmes en cours d'exécution.

Forces et faiblesses de la recherche scientifique française

Forces La recherche française, notamment à travers la communauté SoC², dispose d'un spectre thématique large allant au sens des sections CNU de la section 27 à la section 63. Ceci permet une richesse dans l'approche collaborative, construite autour d'une communauté bien structurée.	Faiblesses La dimension pluridisciplinaire de l'IA innovante s'inscrit à l'interface de disciplines en dehors du champ thématique du SoC², telles que les neurosciences ou la physique fondamentales voire la sociologie. Le potentiel d'innovation de rupture s'entend clairement dans ce contexte, il est donc essentiel de structurer via des moyens incitatifs thématiquement ouverts à une communauté interdisciplinaire.
Opportunités - Les opportunités essentielles s'entendent thématiquement, dans le défi de l'interdisciplinarité en catalysant des recherches à l'interface avec certaine communauté, en s'appuyant sur le spectre sinon complet du moins large des compétences en IA dans le GdR SoC². Le SoC² est également au cœur des autres défis transverses liés à l'IA telle que les questions de sécurité, certifiabilité qui créent donc autant d'opportunités d'innover. Le contexte actuel du plan de relance avec l'emphase sur l'aspect IA embarqué est également largement propice à mettre en avant des projets ambitieux avec des partenaires industriels. - L'IA embarqué a aussi été identifiée comme primordiale par les industriels français et européens, et notamment au travers de la filière stratégique de l'électronique.	Menaces L'importance que revêt l'IA dans le monde des systèmes embarqués fait que la dynamique industrielle est très significative sur le plusieurs sujets relatifs, comme notamment s'agissant des accélérateurs embarqués. Il est alors crucial de bien définir les sujets et les modalités de collaboration pour être en capacité d'écrire l'état de l'art et non de le paraphraser.

Acteurs académiques français (liste non-exhaustive par ordre alphabétique) :

La quasi-totalité des acteurs du GdR SoC² sont actifs sur des thématiques scientifiques liées à l'IA. Le GdR a mené en 2019 une consultation sur 3 grands axes, vouée à estimer les points saillants de l'activité de la communauté. La consultation a mené au recensement suivant :

- Axe 1: IA, Composants, Technologie et Outils de conception (46% des membres du GdR)
Neuro-inspiré, nouvelles technologies, *IA-driven CAD*, frameworks etc.
- Axe 2: IA, Architectures et Applications (37% des membres du GdR)
Accélérateurs pour l'inférence, IA @ edge etc.
- Axe 3: IA, Sécurité, Fiabilité, Test (17% des membres du GdR)
IA pour attaque, IA & *approximate computing*, IA & Test ...

La dynamique d'évolution des sujets est significative aussi une approche orientée portail de référencement est envisagée, permettant aux membres des communautés (GdR SoC², GdR BioComp, GdR MADICS notamment) de s'inscrire et d'y voire référencés leurs travaux en lien avec SoC². Ceci pourrait contribuer à l'émergence d'une communauté transdisciplinaire disposant d'un potentiel d'impact scientifique notable.

Synthèse des conclusions et recommandations

1. Identification des défis et les pistes de recherche à privilégier

En ce début de décennie, le domaine des system-on-chip, des systèmes embarqués et des objets connectés est face à d'importants défis, pilotés essentiellement par la demande incessante de plus en plus de puissance de calcul et de capacité à traiter les informations, aussi bien au sein des system-on-chip qu'à l'échelle des systèmes interconnectés à grande échelle (IoE – Internet of Everything). En même temps, il y a un véritable renforcement de la prise de conscience des défis comme la consommation des ressources (énergie et matériaux), la fiabilisation et la sécurisation des systèmes et des données. De réelles opportunités émergent, sur le plan scientifique comme l'IA embarquée, mais aussi sur le plan contextuel comme le mouvement Open Hardware et les efforts de souveraineté numérique européenne.

- La **complexité** est toujours croissante ; la réduction des dimensions est toujours de vigueur et la Loi de Moore n'est toujours pas finie
- L'hétérogénéité est croissante également, pour l'augmentation de la fonctionnalité ou pour améliorer les interfaces
- Mais il y a une nécessité de consommer moins de ressources : **énergétiques** d'une part, principalement pour les actions de calcul et de communication ; et de matériaux d'autre part, particulièrement pour les matériaux ayant un impact fort sur l'environnement ou présentant un risque de rupture d'accès pour des raisons géopolitiques
- La course vers les **communications** 6G vont également mobiliser les recherches sur les communications THz et la convergence électronique-optique
- Une opportunité intéressante est l'arrivée à maturité d'une zoologie de technologies **mémoires** – il s'agira d'inventer les architectures permettant de les utiliser au mieux
- Un virage très important concerne l'émergence de **l'IA**, et l'impossibilité actuelle de déployer les modèles complexes d'IA vers les nœuds embarqués contraints en ressources énergétiques et en mémoire. Pouvoir créer des supports matériels pour l'IA embarquée est un enjeu capital
- La course à la réalisation de calculateurs **quantiques** pour des applications réelles est engagée, nécessitant l'exploration de systèmes matériels à fort volume quantique
- Avec plus d'intelligence embarquée, il est aussi nécessaire de garantir plus de fonctions, de **sécurité** et de **fiabilité** dans les systèmes embarqués
- L'un des grands enjeux de prochaines années sera l'avènement des systèmes massivement distribués (IoE), où les **applications** diversifiées vont nécessiter des approches spécifiques
- Pour améliorer l'efficacité de conception, le mouvement "**open hardware**" est une réelle opportunité ainsi qu'un enjeu majeur pour la décennie à venir
- Enfin, la reconnaissance de l'importance stratégique et critique de la filière électronique aux niveaux national et européen conduit à des efforts de **souveraineté numérique européenne** pour lesquels la communauté de recherche sera force de proposition

Il est à noter que ces défis sont pour la plupart interdisciplinaires voire transdisciplinaire et mettent en jeux des compétences larges et variées (de la micro-électronique à l'informatique embarquée) mais nécessiteront aussi des ouvertures vers la physique ou les sciences humaines.

Dans les paragraphes qui suivent, nous identifions les défis et les pistes prometteuses de recherche pour répondre à ces défis.

a. Des contextes à forts enjeux – évolutions attendues et pistes de recherche

i. Technology push – Scaling and complexity

Malgré des coûts de fabrication très élevés (à titre d'exemple, les coûts de la fonderie TSMC Fab18 en Arizona aux États-Unis, qui cible la technologie 5nm - actuellement le processus le plus avancé en

production – sont estimés à \$12Md), la Loi de Moore continue. Cette loi empirique, illustrée dans la figure ci-dessous, mesure l'évolution du nombre de transistors par puce, directement reliée à la puissance de calcul des processeurs.

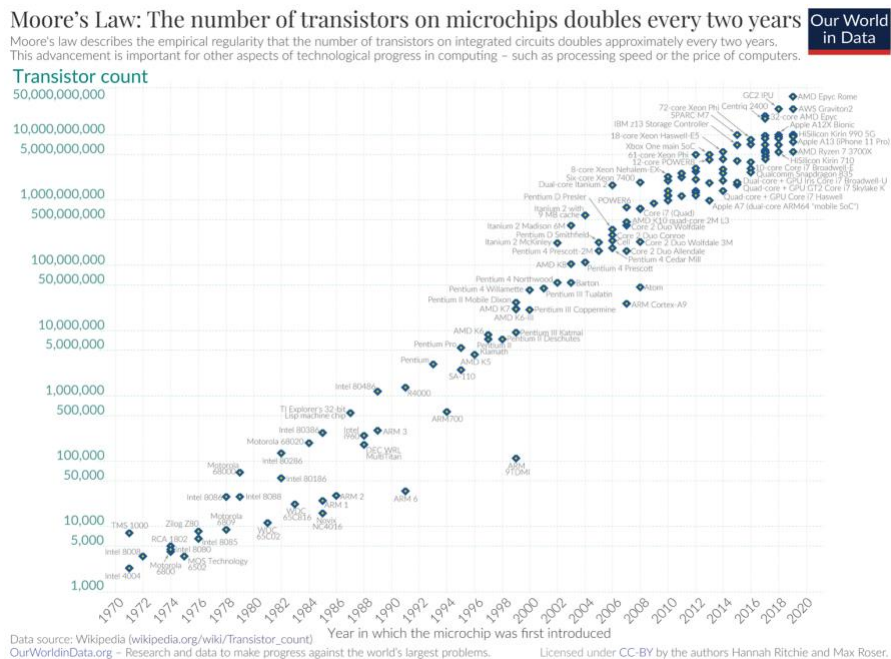


Fig. 1 : Nombre de transistors par puce en fonction du temps (source: [Max Roser, Hannah Ritchie, Our World in Data](#))

L'augmentation du nombre de composants élémentaires ("More Moore") à disposition pour la réalisation de systèmes de calcul a conduit à l'émergence d'architectures manycore (multi-cœur) pour paralléliser les calculs ou pour permettre le traitement parallèle de plusieurs tâches de calcul, puis à la montée en puissance des architectures impliquant des cœurs hétérogènes et l'augmentation de leur taille et complexité.

À titre d'exemples :

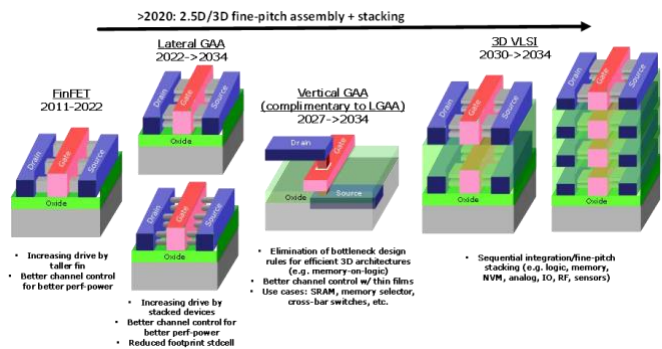
- Le plus grand nombre de transistors dans un **microprocesseur** disponible dans le commerce est de 39,54 milliards de MOSFET, dans le Epyc Rome (2019) basé sur l'architecture Zen 2 d'AMD. Il s'agit d'un circuit intégré 3D (avec 8 puces dans un seul boîtier et 64 cœurs au total) fabriqué en utilisant le processus de fabrication FinFET 7nm de TSMC.
- L'unité de traitement graphique (**GPU**) contenant le nombre de transistors le plus élevé est le GA100 Ampère (2020) de Nvidia avec 54 milliards de MOSFET et 6912 cœurs CUDA, utilisant le processus FinFET 7nm de TSMC.
- Le circuit contenant le plus grand nombre de transistors, **tous circuits intégrés confondus**, est la puce mémoire flash V-NAND (2019) intégrée de 1To de stockage flash universel (empilé 3D) de Samsung, avec 2000 milliards de MOSFET à grille flottante (4 bits par transistor). Un cas à part est le moteur Deep Learning Wafer Scale Engine 2 (2020) de Cerebras, qui est constitué d'un **wafer complet de puces interconnectées**. Il dispose de 2600 milliards de MOSFET, fabriqués en utilisant le processus FinFET 7nm de TSMC.

Alors que la réduction des dimensions des transistors était auparavant le seul vecteur permettant d'augmenter le nombre de transistors par puce, aujourd'hui cette augmentation s'appuie également

sur l'empilement 3D. Cette tendance s'accroîtra dans les années à venir, comme indiqué dans le tableau ci-dessous extrait de l'IRDS 2020. En effet, les codes GxxMxxTxx font référence à : G = dimension de la grille du transistor ; M = dimension du pitch minimal métal-métal ; T = *nombre de couches de transistors empilés*. Ainsi l'augmentation de la complexité à l'horizon 2030 dépend, en plus de changements profonds d'architecture transistor des FinFETs planaires 2D d'aujourd'hui aux structures à grille entourant (*Gate-All-Around*), de l'empilement 3D monolithique.

YEAR OF PRODUCTION	2020	2022	2025	2028	2031	2034
Logic industry "Node Range" labeling (nm)	G48M36	G45M24	G42M20	G40M16	G38M16T2	G38M16T4
IDM-Foundry node labeling	i7-f5	i5-f3	i3-f2.1	i2.1-f1.5	i1.5e-f1.0e	i1.0e-f0.7e
Logic device structure options	FinFET	FinFET	LGAA	LGAA	LGAA-3D	LGAA-3D
Mainstream device for logic	FinFET	FinFET	LGAA	LGAA	LGAA-3D	LGAA-3D
LOGIC TECHNOLOGY ANCHORS						
Patterning technology inflection for Mx interconnect	193i, EUV DP	193i, EUV DP	193i, EUV DP	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV
Beyond CMOS as complementary to mainstream CMOS	-	-	-	2D Device, FeFET	2D Device, FeFET	2D Device, FeFET
Channel material technology inflection	SiGe25%	SiGe50%	SiGe50%	Ge, 2D Mat	Ge, 2D Mat	Ge, 2D Mat
Process, technology inflection	Conformal doping, Contact	Channel RMG	Lateral/Atomic Etch	Non-Cu Mx	3D VLSI	3D VLSI
Stacking generation inflection	2D	3D stacking: W2W, D2W Mem-on-Logic	3D stacking: W2W, D2W Mem-on-Logic	3D stacking: Fine-pitch stacking, P-over-N, Mem-on-Logic	3D stacking: 3D VLSI: Mem-on-Logic with Interconnect	3D stacking: 3D VLSI: Logic-on-Logic

(a)



(b)

Fig. 2 : Prévisions d'évolution dans les technologies de fabrication de circuits intégrés a) Feuille de route portant sur les dimensions b) Architectures futures des transistors élémentaires. Source : "Executive Summary," [International Roadmap for Devices and Systems, IEEE, 2020](#)

Même si le développement technologique pour la mise à l'échelle est essentiellement hors Europe (cette situation peut évoluer, cf. paragraphe 1.a.vi Souveraineté européenne en microélectronique), la communauté française de recherche est toujours à la pointe concernant les problématiques de conception et d'utilisation des architectures complexes de calcul. Une problématique supplémentaire (et notamment identifiée par les industriels) concerne l'intégration 2.5D, le packaging et la conception de SiP³⁰.

Défis	Pistes de recherche
Passage à l'échelle des méthodologies de conception	Abstraction des architectures par des techniques de virtualisation, de synthèse de haut-niveau et d'intégration des systèmes hétérogènes
Approche holistique de conception	Solutions basées sur des accélérateurs matériels, émulateurs, ou " <i>hardware in the loop</i> "
Vérification	Nouvelles méthodes de vérification formelle (preuve)
Résilience des systèmes complexes (à des actes malveillants ou aux perturbations environnementales)	Capacités d'autodiagnostic et d'autoréparation des systèmes

ii. Technology push – hétérogénéité

Les recherches portant sur la co-intégration et l'intégration monolithique de technologies hétérogènes ("*More than Moore*") de type MEMS, optique, fluide, thermique ..., amorcées il y a 30 ans et portées principalement par l'Europe depuis, permettent aujourd'hui :

³⁰ System in Package

- De coupler le monde numérique au monde physique avec une interaction cyber-physique plus dense et/ou avec une efficacité énergétique accrue
- D'augmenter la fonctionnalité des technologies CMOS à venir ou actuelles

Interaction cyber-physique

En raison de la nature analogique (temps continu) des signaux en jeu et la sensibilité des informations aux bruits et à la variabilité, les technologies sub-decananométriques sont peu adaptées aux applications d'acquisition. Les circuits d'interface (analogique / mixte / RF) s'appuient ainsi pour la majeure partie sur des technologies électroniques matures (actuellement FDSOI 28nm et au-delà). Leur évolution est donc pilotée par les systèmes et les applications, plutôt que par les technologies.

Défis	Pistes de recherche
Amélioration des mesures (résolution, bande passante, consommation énergétique)	Asservissement / compensation électronique spécifique, utilisation de l'IA
Réduction du volume de données acquises à transmettre, extraction de l'information pertinente localement	Compression locale des données (Analog to Information), extraction de features (Analog to Feature)
Flexibilité de prétraitement, adaptabilité à l'application et à l'environnement de déploiement, prise en compte du packaging	Approches analogiques reconfigurables, utilisation de l'IA pour optimiser les paramètres des circuits
Santé et bien-être numérique – surveillance continue, électrostimulation, interface neuronale directe (BMI – Brain Machine Interface), communication intracorporelle	Interfaçage intégré de composants et de systèmes électroniques avec des systèmes biologiques (biocompatibilité des matériaux, des rayonnements, des températures), miniaturisation des dispositifs biomédicaux
Systèmes multi-capteurs	Fusion des données
Interfaces CMOS ultime et "Beyond-CMOS"	Exploration et démonstration matérielle des circuits d'interface s'appuyant sur des technologies nanoélectroniques très avancées ou sur les technologies s'appuyant sur des variables non-conventionnelles

Fonctionnalité augmentée

L'utilisation de composants hétérogènes comme "*technology booster*" permet d'augmenter la fonctionnalité des technologies CMOS à venir ou actuelles. Il s'agit notamment de deux technologies de pointe :

- Photonique silicium – Les dispositifs de photonique silicium, fabriqués à l'aide des techniques de fabrication de semi-conducteurs existantes, permettent la création de dispositifs hybrides dans lesquels des composants optiques et électroniques sont intégrés sur une seule puce. La photonique silicium peut être considérée comme un moyen de poursuivre la loi de Moore, en utilisant des interconnexions optiques rendant possible le transfert de données plus rapide inter- et intra-puce. Cette technologie s'est également révélée prometteuse pour des fonctions de capteur, tels que les gyromètres utilisant l'interférométrie par exemple. Plus récemment, en conjonction avec des dispositifs de mémoire non volatile, son potentiel pour la réalisation d'accélérateurs IA a été démontré avec de nouveaux paradigmes de calcul tels que le *reservoir*

computing (type de réseau neuronal récurrent qui projette les signaux d'entrée dans des espaces de calcul de plus grande dimension grâce à la dynamique d'un système fixe et non-linéaire appelé réservoir), les réseaux de neurones profonds et le calcul stochastique.

- Mémoires non-volatiles – l'émergence de nouvelles technologies mémoire à l'échelle nanométrique (et l'arrivée à maturité pour certaines d'entre elles) permet de repenser la hiérarchie classique de mémorisation voire de remplacer la SRAM, la DRAM et la mémoire Flash traditionnelles dans les applications appropriées, par exemple en utilisant des mémoires compatibles CMOS telles que la RAM à changement de phase (PCRAM), la RAM résistive (ReRAM), la RAM magnétique (MRAM) ou les mémoires ferroélectriques (FeRAM ou FeFET). Plus de détails sur la façon dont ces dispositifs peuvent être utilisés pour surmonter les goulots d'étranglement actuels des architectures de calcul sont décrits dans la section sur les Mémoires – Nouvelles technologies, nouveaux usages, nouveaux paradigmes.

La mise à l'échelle de la technologie de traitement de l'information, bien au-delà de ce qui peut être atteint par le CMOS ultime ("Beyond CMOS"), nécessitera

- De nouveaux **paradigmes de calcul** tels que le *calcul neuromorphique* ou le *calcul quantique*,
- De nouvelles **architectures**,
- Des percées technologiques au niveau des **dispositifs** utilisant
 - Des **charges** (par exemple, des transistors à faible pente sous le seuil), ou
 - (à plus long terme) des **variables d'état hybrides ou alternatives** (par exemple spin, magnon, phonon, photon, électron-phonon, photon-supraconducteur qubit, photon-magnon) ; les états étant binaires (numériques), multi-niveaux, analogiques ou intriqués.

Défis	Pistes de recherche
Efficacité (débit, énergie / bit) des interconnexions intra- ou inter-processeur	Réseaux photoniques / RF sur puce, réseaux 3D
Accélération et efficacité du traitement de masses de données	In-network photonic computing (DNN), calcul stochastique photonique, calcul hyperdimensionnel, arithmétique non-conventionnelle
Surmonter le mur de la mémoire	New memory-centric compute paradigms based on emerging memory technologies De nouveaux paradigmes de calcul centrés sur les données et s'appuyant sur les technologies mémoire émergentes
Permettre le calcul " <i>normally-off</i> " et " <i>intermittent</i> " pour l'IoE	Exploitation de nouvelles technologies de mémoires non-volatiles
Nouveaux paradigmes de calcul " <i>Beyond-CMOS</i> "	Exploration et démonstration matérielle des circuits et architectures de calcul à variable d'état alternatif Interfaçage des paradigmes émergents (par exemple quantique) avec des systèmes conventionnels

Intégration hétérogène

Dans une première approche, de telles techniques d'interaction cyber-physique et de fonctionnalité augmentée seront nécessairement couplées à des cœurs de calcul CMOS existants et nécessiteront des techniques d'intégration hétérogènes extrêmes.

Il existe une grande variété de techniques d'intégration, à des échelles diverses. Il s'agit de tirer profit de ces techniques – en effet, l'intégration 3D monolithique homogène ou hétérogène au niveau de la puce, ou l'intégration 2.5D hétérogène au niveau du packaging (empilement vertical, *chipselets*) ne sont que des outils à sélectionner au cas par cas pour optimiser les performances du système et minimiser les coûts. Ainsi, cela rajoute un levier supplémentaire (et donc un degré de complexité supplémentaire) à la conception de ces systèmes déjà très complexes.

Méthodes de conception

Il est crucial, avec une telle complexité sans précédent, de maîtriser l'hétérogénéité des systèmes à tous les niveaux : technologique, architectural, et logiciel. Cet objectif, à visée clairement multi-niveau, joue un rôle absolument central dans la définition de solutions. En effet, grâce aux leviers spécifiques offerts par chaque paradigme de calcul, architecture et technologie, diverses problématiques en jeu pour l'avènement de systèmes à la fois performants, efficaces en énergie et prédictibles seront plus aisément adressables.

Défis	Pistes de recherche
Simulation du système complet	Modélisation / émulation haut-niveau permettant une simulation système à grande échelle tout en permettant une précision suffisante tenant compte des principaux paramètres influant sur le système Techniques efficaces de simulation hétérogène (signaux mixtes, technologies et communications émergentes)
Valider les exigences à l'échelle des systèmes cibles	Méthodes pour maîtriser l'intégration/combinaison des composants hétérogènes
Appréhender les compatibilités technologiques à différentes échelles, prise en compte de l'intégration 3D monolithique de technologies hétérogènes ou de l'empilement hétérogène des puces	Nouveaux outils de simulation multi-physique portant sur les niveaux des procédés de fabrication, des composants, des circuits et des systèmes
Identification des technologies (composants, intégration) les plus prometteuses	Outils de <i>pathfinding</i> (DTCO ³¹ / STCO ³²), optimisation multi-niveau

iii. Application pull – Internet of Everything (IoE)

L'Internet of Everything (IoE) est "la connexion intelligente des personnes, des processus, des données et des objets" et étend le concept de l'Internet des objets (IoT) qui porte sur les communications de machine à machine (M2M) pour décrire un système plus complexe qui englobe également les interactions de machine à personne (M2P) et de personne à personne assistées par la technologie (P2P). L'IoE comprend ainsi les communications et les interactions générées par l'utilisateur associées à l'ensemble global des appareils en réseau, ainsi qu'à tout type d'objet ou d'entité physique ou virtuelle qui peut être rendu adressable et doté de la capacité de transmettre des données.

³¹ Design-Technology Co-Optimization

³² System-Technology Co-Optimization

Le déploiement d'objets connectés à grande échelle et leur utilisation à l'échelle de grands systèmes tels que l'IoE conduira à un développement spécifique par application et selon les usages induits. La figure ci-dessous illustre les domaines applicatifs où l'IoT (et a fortiori l'IoE) est utilisé.

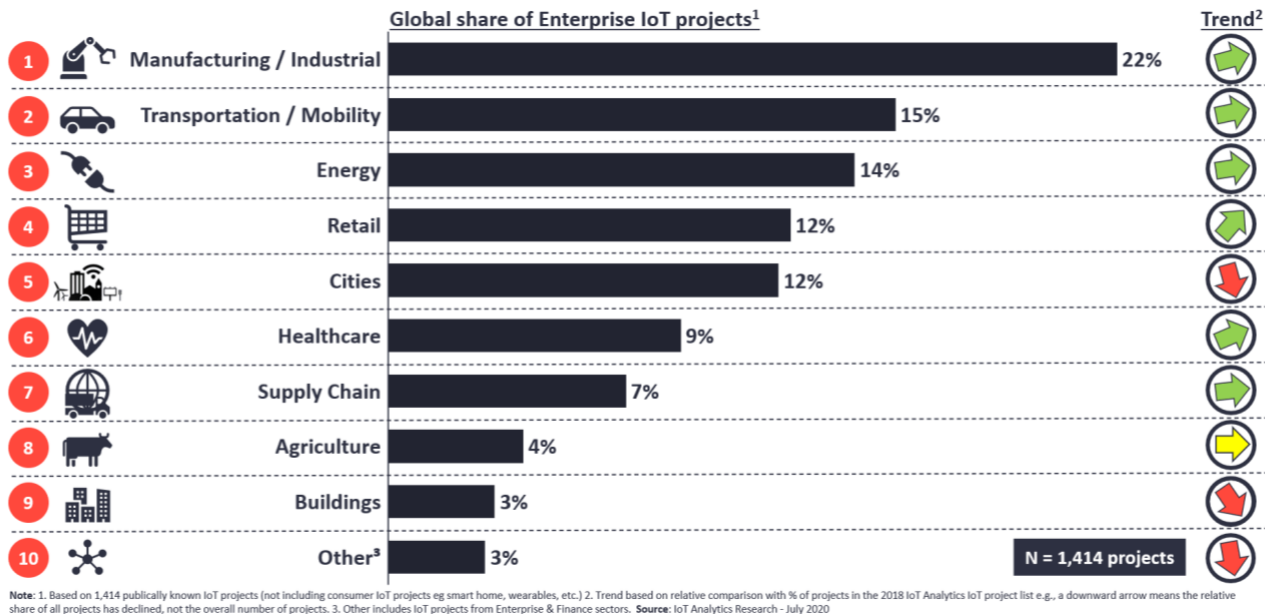


Fig. 3 : Domaines applicatifs 2020 de l'IoT. Source : [IoT Analytics](#).

Défis	Pistes de recherche
Fiabilité, protection et résilience pour applications numériques critiques (santé, aéronautique, défense, transport autonome ...)	Sécurité et intégrité matérielle des fonctions assurées par les objets connectés, des sources et systèmes de conversion d'énergie, et des communications RF
Applications transport : latence réduite et critique (véhicule à véhicule – V2V)	Traitements de données en périphérie (<i>edge computing</i> , <i>near-sensor computing</i>)
Applications santé : privauté des données patient, confiance du couple patient-médecin (surveillance des paramètres physiologiques et/ou activation de la libération de molécules dans le cadre d'un traitement à distance)	Sécurité des données captées puis transmises Explicabilité et fiabilité Interfaces et systèmes biocompatibles Consommation sans pile BAN ³³
Applications Industrie 4.0 : réduction de trafic sur les réseaux pour assurer une densité de connectivité d'un grand nombre d'objets, maintenance prédictive	Convergence des réseaux (5G) Définition de nouveaux protocoles réseaux Jumeaux numériques

³³ Body Area Network

Dispositifs pourvus de capacités d'apprentissage, traitement des masses de données	Modèles d'implémentation efficace du calcul d'inférence, réduction de la précision de calcul tout en conservant une qualité de convergence suffisante IA distribuée
Déploiement grande échelle (millions), hyper-connectivité	Apprentissage distribué, modèles hiérarchiques Coopération efficace entre le <i>Edge Computing</i> et le <i>Cloud</i> , étude du compromis entre approches distribuées et centralisées (données, placement des ressources de calcul et de contrôle, fiabilité et consommation énergétique globales) Simplification de l'exploitation au niveau applicatif
Étudier les comportements (dans toutes ses dimensions) d'un réseau d'objets connectés, face à différents scénarios	Jumeau numérique d'un système IoE, de ses acteurs et de son environnement

iv. Matériel open-source

L'esprit "*open source*", déjà bien établi dans le domaine du logiciel, s'est étendu au matériel avec des systèmes embarqués accessibles tels que Arduino et Raspberry Pi, et a dynamisé une nouvelle communauté de "*makers*" qui construisent des systèmes embarqués au niveau des circuits imprimés. Les jeux d'instructions open-source tels que RISC-V avec un écosystème matériel et logiciel permettent la liberté de conception de systèmes de calcul industriels et consolident la base stratégique des semi-conducteurs. Ce nouvel écosystème autour du RISC-V permet aussi de récupérer de la souveraineté dans les systèmes électroniques et constitue un atout fort pour la France qui possède encore des centrales technologiques. C'est un point fort qu'il conviendra de pérenniser voire améliorer.

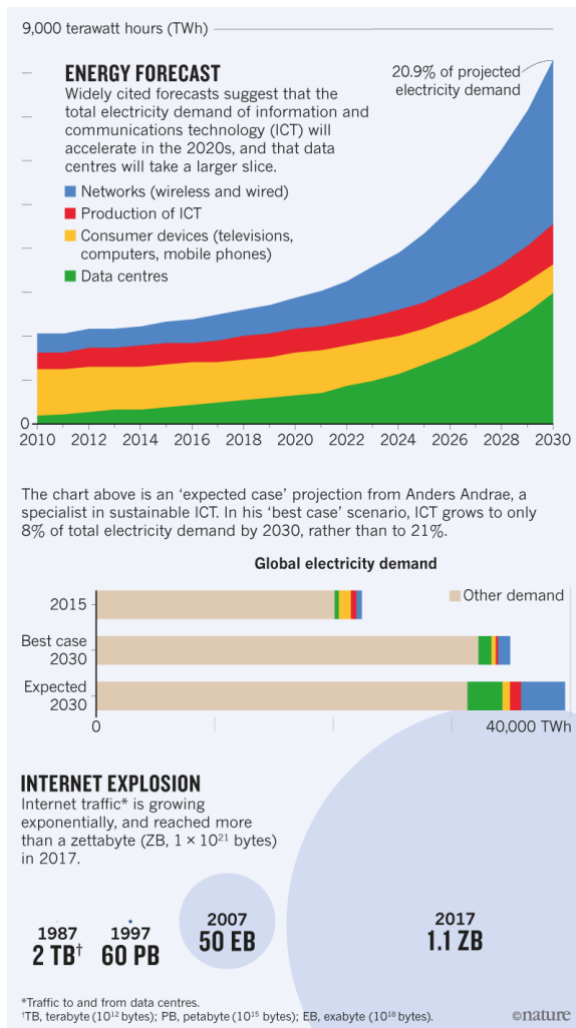
D'autres développements révolutionnaires ont également ouvert la porte à la démocratisation de la conception de circuits intégrés. En 2020, Google, SkyWater et eFabless se sont associés pour lancer un programme basé sur le processus open source SKY130 de SkyWater (CMOS 130 nm). Cette technologie ouverte est proposée à la communauté avec un flot de conception complet pour permettre aux concepteurs de mettre en œuvre leurs idées. Ces initiatives sont d'une importance cruciale pour (i) les normes, (ii) augmenter le niveau de conception (c'est-à-dire se concentrer sur des préoccupations plus élevées au niveau système), (iii) assurer la fiabilité et la sécurité, et (iv) le renforcement de la communauté.

Défis	Pistes de recherche
Accélérateurs matériels RISC-V	Accélérateurs IA dédiés - neuromorphique, approximatif, stochastique, quantique, hyper-dimensionnel, arithmétique nouvelle ...
Extension au jeu d'instruction	Calcul proche ou en mémoire, sécurité-fiabilité
Interfaces matériel / logiciel	Normalisation Étendre l'open-source aux autres constituants (interfaces, mémoires) des systèmes
Performances du matériel Open-Source	Maîtriser l'environnement Open-Source, approches d'accélérateurs

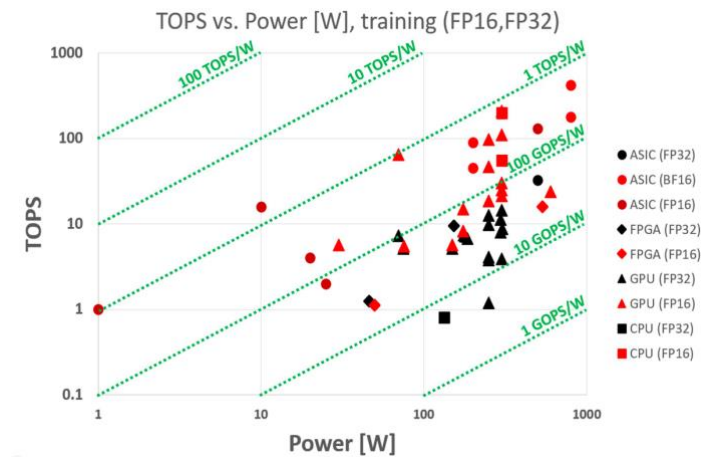
v. Consommation énergétique ICT et développement durable

Aujourd'hui, selon le [Shift Project](#), les technologies numériques (ICT – Information and Communication Technologies) émettent 4% des émissions de gaz à effet de serre, et la part ICT de la consommation d'énergie augmente de 9% par an. La figure ci-dessous représente cette évolution et notamment la prépondérance du calcul (*data centers*) et de la communication (réseaux) dans la consommation énergétique ICT.

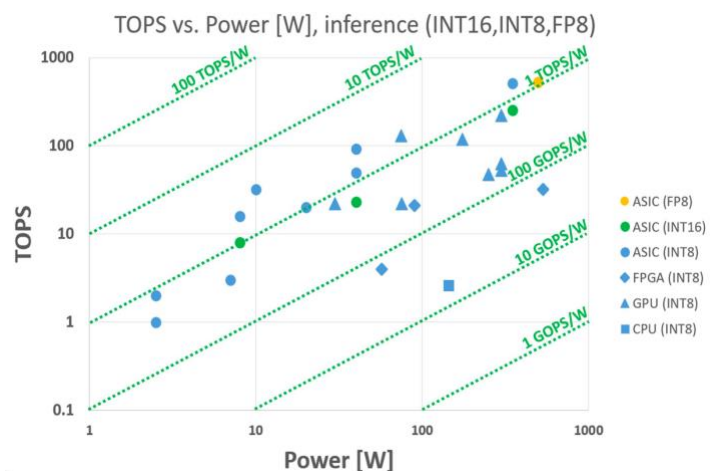
En effet, l'objectif politique actuel de l'UE de réaliser 30% d'économies d'énergie pour minimiser l'impact environnemental de ces systèmes passe nécessairement par l'utilisation de solutions innovantes basées sur les composants électroniques et les systèmes. Enfin, le paradoxe de Jevons énonce qu'à mesure que les améliorations technologiques augmentent l'efficacité avec laquelle une ressource est employée, la consommation totale de cette ressource peut augmenter au lieu de diminuer. Il est donc nécessaire d'accompagner les efforts sur l'amélioration de l'efficacité énergétique par une réflexion sur les usages des technologies numériques.



(a)



(b)



(c)

Fig. 4 : a) Prévisions énergétiques de la consommation ICT. Source : N. Jones, "[How to stop data centres from gobbling up the world's electricity](#)," Nature, Septembre 2018. b) Matériel IA pour l'apprentissage. c) Matériel IA pour l'inférence. Source : "Application Benchmarking," [International Roadmap for Devices and Systems](#), IEEE, 2020.

Pour le développement durable, face à l'explosion du nombre de composants et à la consommation de matériaux rares, il devient urgent de prendre en considération l'utilisation de matières premières et d'énergie à la fabrication du système, ainsi que le coût environnemental du recyclage en fin de vie du système.

L'efficacité énergétique est ainsi une métrique, illustrée par la figure ci-dessus pour le matériel réalisant des calculs d'apprentissage ou d'inférence dans le domaine de l'IA (où TOPS/W signifie 10^{12} opérations/Joule). Il est intéressant de noter que l'efficacité énergétique s'améliore avec la spécificité du matériel (ASIC, FPGA, GPU, CPU). Au niveau système, il est souhaitable de gagner un facteur 10 sur cette efficacité énergétique dans la prochaine décennie pour l'ensemble des systèmes numériques.

Néanmoins, l'évolution actuelle de traitement de données en périphérie (*edge computing*, *near-sensor computing*, *edge AI*) s'accompagne de problématiques inédites en termes de capacité de calcul embarquée, sous contrainte de ressources (principalement énergie, mais aussi mémoire), et également de prédictibilité pour répondre au besoin des fonctions temps-réel :

- Rendre possible en embarqué des tâches jusqu'ici cantonnées aux systèmes conventionnels
- Réduire les besoins en termes de production et de stockage d'énergie.

Défis	Pistes de recherche
Augmenter l'efficacité énergétique globale	Poursuivre, exploiter et combiner les recherches sur les leviers de faible consommation à tous les niveaux (technologies, composants, circuits, architectures, logiciel, OS, application) Réflexion sur les usages
Calcul éco-énergétique	Méthodes pour adapter le calcul aux besoins (calcul approximé), accélérateurs matériels, méthodes centrées sur l'énergie ou sur la mémoire Optimisation des hyper-paramètres des réseaux, représentation efficace des données, nouveaux substrats de calcul
IoE éco-énergétique	Traitements de données en périphérie (<i>edge computing</i> , <i>near-sensor computing</i> , <i>edge AI</i>) pour réduire le volume des communications, gestion flexible des phases d'activité / d'inactivité du capteur, efficacité énergétique de la transmission RF
Identification des technologies (composants, intégration) les plus prometteuses	Mémoires non-volatiles, energy-harvesting, battery-less systems
Durabilité et réutilisabilité accrue des systèmes	Utilisation de composants matériels et logiciels libres Réutilisation (via la reconfiguration par exemple) Réflexion sur les usages

vi. Souveraineté européenne en microélectronique

L'industrie des semi-conducteurs et des circuits intégrés bénéficie aujourd'hui d'une chaîne logistique internationale bien huilée. Cependant, les événements récents (pandémie COVID, guerres commerciales) mettent en évidence sa fragilité – ainsi que sa criticité pour l'offre de circuits intégrés et par conséquent pour l'économie. Conformément aux efforts actuels pour revitaliser le paysage

industriel européen, l'UE a annoncé début 2021 des plans à long terme pour des processus avancés des semi-conducteurs et des moyens de fabriquer des circuits intégrés avancés, renversant ainsi une tendance de 30 ans d'externalisation progressive vers des pays d'autres continents (en particulier aux États-Unis et à Taiwan). En effet, la concentration des ressources sur quelques régions géographiques seulement (cf. figure ci-dessous) signifie que toute perturbation dans une région peut affaiblir ou même briser la chaîne d'approvisionnement dans son ensemble (en témoignent les difficultés d'approvisionnement actuelles en circuits intégrés).

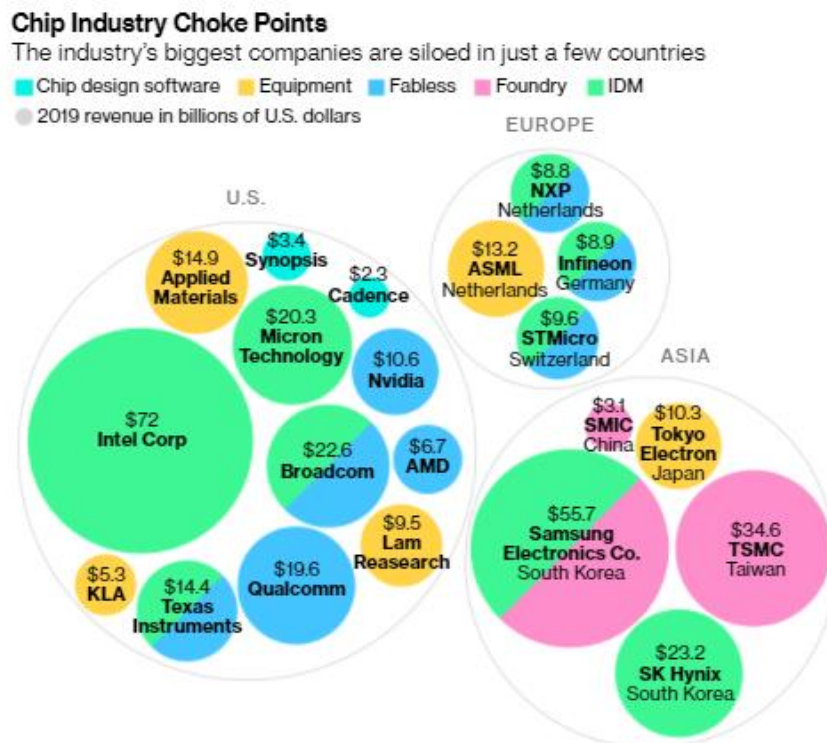


Fig. 5 : Goulots d'étranglement dans l'industrie des semi-conducteurs, concentration géographique des ressources et compétences. Source : N. Drozdiak, H. Fouquet, "[Europe Weighs Semiconductor Foundry to Fix Supply Chain Risk](#)," Bloomberg, February 2021

Autrefois centre d'usines de semi-conducteurs, l'Europe a considérablement réduit le niveau de fabrication au cours des 20 dernières années, notamment car les concepteurs de circuits intégrés automobiles (NXP, Infineon) préfèrent externaliser une grande partie de la production à des géants comme TSMC et autres fonderies. Aujourd'hui, les constructeurs automobiles en particulier sont pénalisés par des pénuries de semi-conducteurs – Volkswagen et Renault ont perdu des dizaines de milliers de voitures en production par exemple. Bien que le problème puisse s'avérer de courte durée, le problème a mis en évidence la dépendance de l'Europe à l'approvisionnement en technologies clés à l'étranger.

L'Europe représente actuellement moins de 10% de la production mondiale de processeurs et d'autres produits microélectroniques. Par conséquent, et pour renforcer sa position dans ce secteur critique et stratégique, l'UE a défini l'année dernière un objectif : produire au niveau mondial au moins 20% des circuits intégrés et microprocesseurs en valeur. En particulier, elle construira une fonderie avancée en Europe dans le but de réduire la dépendance vis-à-vis des États-Unis et de l'Asie pour des technologies telles que les puces destinées à alimenter les systèmes sans fil 5G, les voitures

connectées, le calcul haute performance, etc. L'objectif est d'explorer comment produire des circuits intégrés avec des caractéristiques inférieures à 10nm, et ce jusqu'à 2nm.

Cette initiative reconnaît que la souveraineté numérique européenne dépend d'une autonomie européenne en microélectronique. Bien qu'elle se concentre principalement sur les principaux fabricants européens et les grands acteurs de la recherche, la communauté académique a également un rôle à jouer pour résoudre de nombreux défis liés à la construction de puces fiables, sûres, éco-énergétiques et durables avec un niveau de complexité aussi élevé, et en focalisant les applications de l'IA embarquée et l'IoE. L'étude des aspects packaging, abandonnée il y a plusieurs années, reprends aussi de l'intérêt (et notamment sur le plan industriel) dans le cadre de la fabrication de systèmes complexes et hétérogènes comme évoqué ci-dessus.

vii. Intégrité et fiabilité

Les processeurs font aujourd'hui appel à plus de composants élémentaires, individuellement plus difficiles à maîtriser sur le plan technologique. Avec cette augmentation continue des densités et des complexités d'intégration, le problème du test de circuits et systèmes intégrés est devenu beaucoup plus pointu, et nécessite de nouvelles solutions économiquement viables en termes de coût et d'efficacité du test. Outre ce problème lié à la complexité de croissance incessante des circuits et systèmes intégrés, d'autres problèmes apparaissent désormais et rendent le test exponentiellement plus difficile :

- La variété des modes de défaillance selon les types de circuit (numérique, analogique, mixte, RF, etc.) et les processus technologiques (2D, 2,5D, empilement 3D) ;
- La variété de nouveaux composants rendus possibles par l'intégration hétérogène – en particulier les technologies de mémoire non volatile comme les mémoires résistives (ReRAM) ou magnétiques (STT et SOT MRAM) ;
- Les nouveaux paradigmes de calcul, en particulier le calcul neuro-inspiré et le calcul approximatif, intrinsèquement non déterministes, ainsi que le calcul en mémoire ;
- De nouvelles opportunités pour tester les circuits complexes en s'appuyant sur les méthodes basées sur des techniques IA
- Les analyses de défaillances complexes et extrêmement coûteuses, qui imposent des stratégies de diagnostic de défauts plus performantes, pour in fine améliorer les rendements de production.

Par ailleurs, les problèmes de fiabilité dans les processeurs sont exacerbés en raison des éléments suivants :

- Technologies de plus en plus profondes et de plus en plus sensibles aux perturbations environnementales (rayonnements électromagnétiques en particulier) ;
- Besoin de calculateurs fiables pour de nombreuses applications critiques (automobile, santé, aéronautique et spatial, etc.) ;
- Apparition de nouveaux paradigmes (calculateur quantique).

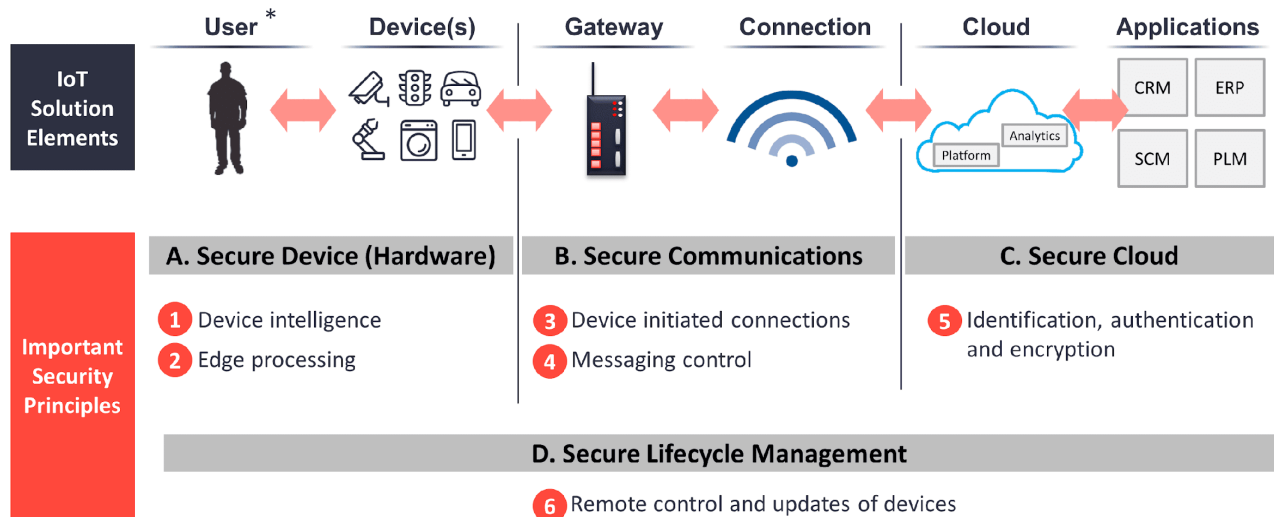
Défis	Pistes de recherche
Qualité des tests en termes de couverture de défauts	Nouvelles approches (ATPG, DFT, BIST) ciblant les défauts au niveau transistor
Diagnostic et analyse de défaillances	Utilisation de techniques d'apprentissage
Efficacité des tests pour les circuits analogiques et RF	Utilisation de techniques d'apprentissage
Test des architectures basées sur les nouveaux paradigmes de calcul	Nouvelles métriques d'évaluation de la qualité des tests

Fiabilité et robustesse des dispositifs élémentaires	Systèmes d'autotest, d'autodiagnostic et d'autoréparation embarqués, systèmes tolérants aux fautes – vers la "dependability" des systèmes
Fiabilité des calculateurs quantiques	Nouvelles approches basées sur la redondance

viii. Sécurité et privauté

La sécurité doit faire partie intégrante de la chaîne de valeur de l'IoT / IoE, depuis l'utilisateur ou le système en périphérie, jusqu'au matériel et aux applications fonctionnant dans le cloud ; en passant par plusieurs couches de liaisons de communication. Cela implique une infrastructure permettant de sécuriser l'ensemble incluant notamment la génération de fonctions physiquement non-clonables, des drivers et circuits sécurisés à différents niveaux, des systèmes IoT / IoE auto-configurables ; le tout contribuant à assurer la sécurité de l'infrastructure et de toutes les interactions entre des millions de nœuds.

Sur le plan de la privauté, la création d'une *société numérique ouverte et citoyenne* nécessitera de savoir construire des systèmes respectueux de la vie privée, à même d'être inscrits dans des plans de gouvernances des données.



Source: IoT Analytics

* User: can represent a person, device, system, or application

Insights that empower you to understand IoT markets

Fig. 6 : Les six principes de sécurité IoT / IoE à tous les niveaux. Source : [IoT Analytics](https://www.iot-analytics.com).

Défis	Pistes de recherche
Améliorer le respect de la vie privée	Traitement local des données brutes en périphérie (<i>edge computing, near-sensor computing</i>) Cryptographie, chiffrements homomorphe embarqué
Améliorer la sécurité par rapport aux chemins d'attaques complexes (mixte, ou à la frontière entre le logiciel et le matériel)	Protections dédiées efficaces exploitant conjointement des niveaux cryptographiques, logiciels, matériels et physiques Sécurité matérielle

Améliorer la sécurité par rapport aux attaques logicielles et physiques (injection de fautes multi-spot, rayons X, analyse de canaux cachés)	Compréhension fine des phénomènes physiques Transistors résistants aux perturbations, écrans protecteurs efficaces et peu coûteux, détecteurs Architectures de processeurs sécurisés par conception
Améliorer la sécurité par rapport aux attaques utilisant les outils et méthodes de Machine Learning / IA	Contre-mesures spécifiques à ces nouvelles attaques
Améliorer la qualité et l'efficacité de conception de systèmes sécurisés	Implémentation (logicielle et matérielle) efficace d'algorithmes de cryptographie Méthodes de conception intégrant les aspects de sécurité, génération automatique de composant sécurisé
Qualité des TRNG, PUF	Utilisation de nouvelles technologies
Éviter la propagation d'apprentissages détournés dans l'IoE	Sécurité de l'apprentissage Mécanismes de certification de confiance
Améliorer la fiabilité des systèmes	IA comme outil d'analyse

b. Des enjeux à fortes valeurs stratégiques : les atouts de la France et de l'Europe

i. Communications – vers la 6G et les THz

Dans les systèmes de communication à l'échelle des systèmes embarqués, l'évolution actuelle prévisible 6G (et au-delà) nécessite d'augmenter les débits tout en réduisant la consommation d'énergie lors des communications. Pour atteindre cet objectif, il est nécessaire d'ici 2030 de réussir la convergence de l'électronique et de l'optique pour permettre la propagation en espace libre des communications optiques. Ces technologies, aux frontières de l'électronique et de l'optique, sont les technologies de communication THz. La connaissance approfondie de leurs caractéristiques (puissance d'émission et de réception, forme d'onde, format de modulation, etc.) permettra d'extrapoler les performances des futurs réseaux locaux.

Défis	Pistes de recherche
Améliorer le rapport débit / consommation des communications	Convergence électronique-photonique – communication THz
Communications RF pour les réseaux 5G-6G	Conception de circuits RF large bande, flexibles et interopérables, sécurisés Gestion avancée des ressources (partage de bandes, ...)
Communications RF pour l'IoT	Conception de circuits RF éco-énergétique, communications durables, IoT coopératif Gestion intelligente et flexible des phases d'activité/d'inactivité de la transmission RF
Convergence calcul-communication	Déploiement de moteurs IA pour traitement partiel dans le réseau <i>edge-cloud</i> (" <i>in-network computing</i> ")

ii. Mémoire : nouvelles technologies, nouveaux usages, nouveaux paradigmes de calcul

Les architectures de calcul conventionnelles souffrent de la latence dans les transferts de données entre des technologies hétérogènes (cœurs de calcul, mémoires volatiles et non volatiles) et la

consommation d'énergie associée au mouvement des données. Le caractère prometteur des nouvelles technologies mémoires (par exemple STT/SOT-MRAM, PCM, ReRAM, FeRAM), et d'interconnexion des composants (par exemple des interconnexions optiques ou l'exploitation de l'empilement 3D – *3D-stacking*) permet d'envisager une remise en cause plus large des architectures des systèmes, au-delà d'une "simple" intégration d'accélérateurs. En effet, l'interaction calcul-mémoire fait l'objet de nombreuses recherches notamment à l'international. L'arrivée concomitante des technologies mémoires non-volatiles rend ces recherches d'autant plus disruptives.

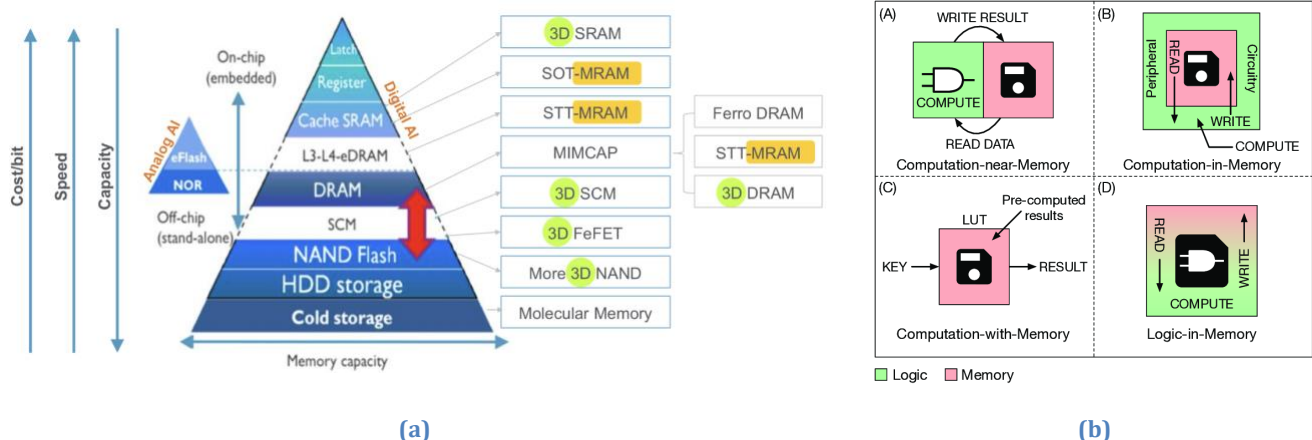


Fig. 7 : a) Hiérarchie mémoire et positionnement des mémoires émergentes. Source : M. LaPedus, "The Next New Memories," *Semiconductor Engineering*, August 15th, 2019. b) Nouveaux paradigmes d'utilisation des technologies mémoire. Source : G. Santoro, G. Turvani, M. Graziano, "New Logic-In-Memory Paradigms: An Architectural and Technological Perspective," *Micromachines*, vol. 10, no. 368, 2019.

Plusieurs moyens de redéfinir les architectures de calcul pour tirer parti de ces nouvelles technologies de mémoire peuvent être identifiés. En particulier, en fonction de l'utilisation de la mémoire pour le calcul, quatre principales approches de "calcul en mémoire" apparaissent :

- (A) Computation-near-Memory (CnM) : les technologies d'intégration 3D permettent de rapprocher le calcul et le stockage en réduisant la longueur des interconnexions. La logique et le stockage sont toujours deux entités distinctes.
- (B) Calcul en mémoire (CiM) : La structure de mémoire standard n'est pas modifiée, tandis que le calcul des données est effectué dans les circuits périphériques.
- (C) Computation-with-Memory (CwM) : La mémoire est utilisée comme tableau pour stocker puis récupérer les résultats pré-calculés ("*result caching*").
- (D) Logic-in-Memory (LiM) : le calcul est effectué directement à l'intérieur de la mémoire en ajoutant une logique simple dans chaque cellule de mémoire.

Défis	Pistes de recherche
Permettre le calcul proche ou en mémoire, systèmes centrés sur les données	Explorations liant les couches technologiques jusqu'aux couches logicielles (voire langages de programmation) pour favoriser les choix technologiques
Réduire l'empreinte des données et augmenter la robustesse du système	Nouveaux paradigmes de calcul (stochastique, approximatif, hyperdimensionnel etc.) tirant le meilleur parti de nouvelles technologies (quantique, mémoire, etc...) <i>Normally-off computing</i>

Limitier le transfert de données critiques en favorisant le traitement local ou dans le réseau	Conception d'architectures neuro-inspirée, calcul adiabatique, <i>in-network computing</i>
---	---

iii. Intelligence artificielle embarquée

Le développement de l'intelligence artificielle pose de nombreux défis quant à son intégration sur des architectures spécialisées. En quelques années seulement, avec l'avènement des architectures d'apprentissage profond et les applications d'assistance personnelle et de moteurs de recherche poussées par les géants américains GAFA, l'IA est passée du statut de curiosité académique à celui d'une véritable force économique et fait des percées profondes dans tous les principaux secteurs d'application. Cela a été rendu possible par :

- D'une part, le succès des réseaux de neurones profonds pour "donner un sens" aux données avec des taux d'erreur très faibles – par exemple, la reconnaissance vocale en 2019 a démontré un taux d'erreur de seulement 4%.
- D'autre part, la croissance de la puissance de calcul des architectures, capables aujourd'hui de supporter la quantité très importante de calculs demandée par ces algorithmes.

L'IA domine désormais également les nouveaux systèmes de jeux vidéo et imprègne tous les domaines de l'acquisition de données. En fait, AI / ML évolue rapidement vers la périphérie et donne naissance au "*Edge AI*". Cette évolution est bénéfique à plusieurs niveaux :

- L'acceptabilité de l'IA embarquée sera avantagée par l'utilisation d'approches matérielles dédiées pour réduire la consommation énergétique globale d'une part, et par la proximité aux utilisateurs pour démystifier et focaliser sur des applications pour aider les personnes d'autre part.
- La réalisation de solutions IA embarquée peut s'appuyer sur l'écosystème national dynamique dans ce domaine, et contribuera à la relance et à la réindustrialisation nationale et européenne
- Enfin, les besoins de personnel hautement qualifié à tous les niveaux de la chaîne de valeur se font sentir dans cette filière en tension, et l'IA embarquée peut réellement être un vecteur commun de construction

Cependant, les difficultés de réalisation du *edge computing* (cf. section 1.a.v Consommation énergétique ICT et développement durable) ne seront qu'accentuées par la généralisation de l'IA embarquée. Les architectures conventionnelles, centrées sur le calcul ("*Von-Neumann computing*"), atteignent leurs limites pour les applications nécessitant un traitement massif des données. L'exécution efficace des algorithmes issus de l'IA, que ce soit en périphérie ou sur le *cloud*, passe par l'intégration d'accélérateurs spécifiques au sein des architectures. En comparaison des implémentations logicielles sur CPU ou GPU classiques, des implémentations matérielles n'ont plus à démontrer leur efficacité. Il peut s'agir actuellement de (GP)GPU, de TPU, de solutions dédiées à base de FPGA, mais ce ne sera pas suffisant. Un véritable changement de paradigme des architectures des machines est nécessaire, et cela impactera toutes les couches de la chaîne de valeur – des composants aux couches logicielles en passant bien évidemment par la conception de circuits et architectures d'accélérateurs à terme exotiques, y compris quantiques.

Défis	Pistes de recherche
Modèles de calcul IA adaptés au déploiement embarqué	Calcul approximé, calcul en mémoire, calcul neuromorphique
Modèles de calcul IA pour déploiement sur un substrat en rupture	Calcul quantique

Approches IA en rupture (orientée algèbre linéaire)	Réseaux à spikes (SNN), <i>Energy Based Models</i> (EBM)
Portage de moteurs IA sur un support embarqué	Algorithmes, adéquation algorithme / architecture et architectures dédiées pour applications contraintes
Accélérateurs dédiés	Exploration et conception d'accélérateurs dédiés (DNN, RNN, LSTM, arbre flou, réseaux antagonistes génératifs ...)
<i>Edge AI</i> adaptatif	Apprentissage embarqué / incrémental par renforcement
Collaboration humain-machine	Interface IA / humain "augmenté", capteurs intelligents connectés
Modélisation et optimisation de systèmes embarqués complexes – circuits, architectures spécialisées	Méthodes et modèles IA / hybrides en complément des techniques traditionnelles (recherche opérationnelle, statistiques ...)
Méthodes de conception pour IA embarquée, systèmes d'IA peu ou non supervisés, à apprentissage continu	Outils d'optimisation et d'automatisation de réglages de paramètres, d'accélération de l'inférence et de l'apprentissage

iv. Calcul quantique

Il est aujourd'hui admis que le calcul quantique est capable de résoudre rapidement certaines classes de problèmes de calcul, tels que la factorisation d'entiers (opération critique pour l'algorithme de cryptage RSA³⁴), qu'aucune approche conventionnelle de calcul ne pourrait résoudre dans un temps acceptable – il s'agit de la "*suprématie quantique*".

Les faits saillants récents, en particulier aux États-Unis et en Australie, comprennent (i) l'affirmation de Google AI d'avoir effectué, en collaboration avec la NASA, un calcul quantique qui est irréalisable sur n'importe quelle machine de calcul classique, ainsi que (ii) des démonstrations de systèmes à fort volume quantique (mesure des capacités et du taux d'erreur d'un système de calcul quantique, qui dépend du nombre de qubits et de la profondeur du système) accessibles via des interfaces cloud. IBM et Google ont annoncé en 2020 des plans de croissance rapide pour les machines quantiques à 1000 qubits (l'état de l'art est à environ 50-75 qubits aujourd'hui) d'ici 2023 – ce qui serait potentiellement suffisant pour démontrer l'avantage quantique sur une application pertinente. En effet, le calcul quantique poursuivra son développement au cours des prochaines années, notamment avec une évolution vers des cas d'usage concrets tels que l'ingénierie biochimique, la découverte de nouveaux matériaux, l'intelligence artificielle et la sécurité des données.

Pour y arriver, il est essentiel de poursuivre l'amélioration des performances matérielles et la démocratisation de l'accès matériel via de nouveaux services. Les efforts de recherche dans les architectures système, les stacks de programmation et les outils de conception en particulier doivent être intensifiés pour connecter le calcul quantique à des applications réelles. Enfin, étant donné que les qubits sont intrinsèquement fragiles avec une durée de vie limitée, d'autres défis importants concernent (i) la protection des informations quantiques contre les erreurs dues à diverses sources de bruit grâce à une conception physique soignée de circuits d'interface cryogéniques et (ii) la mise en œuvre efficace d'approches de correction d'erreur quantique pour le calcul quantique tolérant aux fautes.

³⁴ Rivest–Shamir–Adleman

Défis	Pistes de recherche
Réalisation d'architectures de calcul quantique complexes	Méthodes et outils de simulation et de conception d'architectures quantiques, articulation avec Qiskit
Améliorer la durée de vie des qubits	Circuits d'interface cryogéniques ultra faible-bruit
Tolérance aux fautes	Approches efficaces de correction d'erreur quantique
Applications réelles	Stacks de programmation / compilation efficaces

2. Recommandations structurelles – Instruments à privilégier pour dynamiser la recherche

La grande diversité des laboratoires nationaux travaillant dans le domaine des system-on-chip, des systèmes embarqués et des objets connectés offre un environnement fertile au développement des recherches. Cette diversité est principalement visible dans les domaines abordés, depuis la modélisation de composants et la conception de circuits jusqu'à la programmation d'architectures de calcul hétérogènes complexes. La communauté SoC² dispose en effet d'un spectre thématique large alliant les sections CNR 27, 61 et 63. Ceci permet une richesse dans l'approche collaborative, construite autour d'une communauté bien structurée construite depuis plusieurs années. Les ruptures technologiques décrites dans la partie précédente nécessitent des compétences larges. Le GdR SoC² a œuvré pour rassembler et rapprocher plusieurs communautés variées possédant ces compétences, et le GdR se positionne naturellement pour relever ces défis.

a. Articulation formation – recherche – industrie

Le domaine représenté par SoC² est d'une importance hautement stratégique pour l'activité économique aux échelles nationale et européenne, mais en France le cercle vertueux formation – recherche – industrie est loin d'être optimal.

Pour l'articulation recherche – industrie, la mise en place d'appels d'offre des industriels, sous l'égide de la DGE, serait souhaitable pour mettre à disposition des académiques les nouvelles technologies afin de faire des preuves de concept de systèmes innovants à TRL bas. Des instruments pour favoriser et accroître les délégations des chercheurs / enseignants-chercheurs dans l'industrie permettraient également de renforcer les liens directs entre les mondes académique et industriel. Ceci est particulièrement important dans le domaine de l'IoT / IoE où les domaines applicatifs seront sans aucun doute des catalyseurs de compétences pour répondre au mieux aux spécificités des usagers. En effet, les domaines du transport intelligent et de la santé, pour ne prendre que ces deux exemples, ne privilégieront pas les mêmes objectifs mais s'appuieront sur des contributions très larges.

Les métiers du secteur électronique sont aujourd'hui en grande tension – l'industrie pointe le manque, entre autres, de personnel technique hautement qualifié. La mise en place du Comité Stratégique de Filière (CSF) Électronique pour redynamiser ce secteur est une opportunité importante. Elle a été saisie par le GdR SoC² à travers son implication, malgré l'absence institutionnelle du CNRS, dans les travaux des GTs Formation et Recherche. Compte tenu de l'absence du CNRS dans ce comité, les orientations de recherche du CSF sont majoritairement industrielles ; néanmoins les laboratoires CNRS se positionnent sur des thématiques proches et en amont. Une autre CSF sur la sécurité est en cours de préparation, et il est important que le GdR SoC² puisse y participer pour ce secteur également très stratégique.

Pour l'articulation formation – recherche, tout d'abord il est urgent d'adapter les programmes de formation pour expliquer l'apport disciplinaire aux enseignements sectoriels et applicatifs, plus lisibles et attractifs. Un point spécifique concerne la nomenclature des mentions Master qui n'intègre

ni les évolutions ni les besoins au niveau de la formation. Il est enfin urgent également d'adresser le problème du manque criant et récurrent d'attractivité de la formation doctorale auprès des étudiants Bac+5 et en particulier pour les jeunes diplômés des écoles d'ingénieurs.

Enfin une recherche amont et en rupture nécessite aussi une certaine liberté scientifique qui nécessite des fonds récurrents stables et suffisants pour le déploiement d'une politique scientifique ambitieuse. Si la recherche par projet permet de soutenir l'innovation, des financements récurrents stables permettront une vision plus long terme sur des thématiques à risque. Cette politique pourrait s'appuyer sur les laboratoires, et les GdR.

b. Favoriser la pluridisciplinarité

Pour parvenir à adresser l'ensemble des défis scientifiques, la complémentarité des laboratoires et des équipes de recherche est indispensable. À cet égard, notons que la structuration de la recherche publique française divisée en sections (CNRS, CNU) n'est pas / n'est plus en adéquation avec l'évolution de la complexité de nos sujets d'étude et introduit même des barrières implicites à des coopérations naturelles.

Un grand nombre de laboratoires français dispose de compétences couvrant tout ou partie des thématiques relevant des objets connectés, mais il n'y a pas, au jour d'aujourd'hui, de pôle d'excellence visible en France alliant des compétences scientifiques complémentaires en électronique, informatique et domaines connexes comme les mathématiques appliquées par exemple.

La dimension pluridisciplinaire de l'IA innovante en particulier s'inscrit à l'interface de disciplines en dehors du champ thématique du SoC², telles que les neurosciences ou la physique fondamentales. Le potentiel d'innovation de rupture s'entend clairement dans ce contexte, il est donc essentiel de structurer via des moyens incitatifs thématiquement ouverts à une communauté interdisciplinaire.

De plus, depuis de nombreuses années, des collaborations nombreuses existent entre les chercheurs des sciences "dures" pour adresser ces challenges, mais le déploiement massif (notamment de l'IoT / IoE et de l'IA) devra prendre en compte des aspects liés aux sciences sociales et humaines. Il sera donc nécessaire d'inclure des collaborations mêlant sciences "dures" et sciences "molles".

En première intention, la mise en œuvre d'instruments structurants comme l'ex-ACI (action concertée incitative) ou de PEPS impliquant un nombre important de partenaires serait une évolution souhaitable, avec une implication de la MITI du CNRS.

c. Passage à l'échelle – une plateforme pérenne multi-partenaire / multi-institut

En raison de la grande pluridisciplinarité des questions de recherche à travailler, et à l'instar de la plateforme Renatech (Réseau national des grandes centrales de technologies), il est souhaitable de mettre en œuvre des réseaux / centres nationaux intégrant toute la chaîne de valeur de conception, (modèles, outils de conception/synthèse et de simulation de circuits et d'architectures, PDK, aspects logiciels, compilateurs, *benchmarking* et *pathfinding*) portant sur :

- L'IoE spécifique aux applications santé, transport, usine du futur
- L'IA allant de la technologie (en lien avec Renatech) jusqu'aux applications (une réflexion est en cours au niveau MESRI)

En effet, Renatech permet la maîtrise et la capacité à fabriquer, caractériser et modéliser les nano-composants à la base de nouvelles architectures système, mais ne permet pas le déploiement de ces nano-composants à l'échelle de circuit au travers de réalisations matérielles ou preuves de concept – ce qui permettrait de calibrer des modèles abstraits de ces circuits pour passer à l'échelle d'architectures complètes. D'autres acteurs à l'échelle mondiale, notamment aux USA, sont performants à ce niveau – et la France a tous les atouts pour le faire également. En effet, les

laboratoires français impliqués dans ces axes de recherche couvrent la quasi-totalité des compétences scientifiques nécessaires au développement de cette chaîne de valeur, à l'exception de l'élément clé permettant le passage à l'échelle du nano-composant unitaire au circuit complexe.

Actuellement, la maîtrise de cet élément clé passe par la mise en place de collaborations ponctuelles, à l'échelle nationale avec le CEA-Léti et à l'échelle européenne avec l'IMEC. Seules ces deux centrales technologiques permettent, en Europe, le passage à l'échelle du nano-composant au circuit complexe en agrégeant des technologies hétérogènes en 3D, à l'instar des centres de recherche multi-laboratoires aux USA ou interuniversitaires en Chine. Pour accroître le succès et la visibilité des nouveaux nano-composants développés dans les laboratoires du CNRS d'une part et d'autre part favoriser la maîtrise des technologies de demain et leur transfert dans l'industrie, la mise en place d'un partenariat institutionnel pérenne entre le CNRS et ces centrales technologiques est indispensable.

Ce besoin est d'autant plus urgent que le développement actuel du mouvement pour le matériel et le logiciel ouvert pour l'architecture en Europe constitue un réel levier. Une implication forte de la recherche française permettrait à la fois de renforcer les ressources pour le développement de la recherche en SoC² et regagner la souveraineté dans ce domaine. Il faut nécessairement passer par une mise en commun et un développement collectif pour espérer rivaliser avec la concurrence mondiale.

d. Moyens et profils de poste à privilégier

Le manque de moyens pérennes, ainsi que la recherche de moyens via les appels à projets ou contrats de recherche dont la complexité demande une montée en compétence qui nécessite du temps, pénalisent le temps utile à la recherche scientifique et technologique dans ce domaine. Le positionnement de financements récurrents permettrait aux unités de recherche de se lancer dans des thématiques de recherche long-terme.

Un investissement national important est souhaitable pour renforcer les ressources humaines des acteurs publics de la recherche de ce domaine. Le tableau ci-dessous indique des propositions de profil pour le recrutement de chercheurs / enseignant-chercheurs dans les prochaines années jugées utiles pour la recherche scientifique française dans le périmètre scientifique couvert par le GdR SoC²:

Section CNRS	Intitulé	Thématiques
07	Architectures et nouveaux paradigmes pour le calcul durable et responsable	Approches sobres et verticales, technologies de sécurité et de privauté, électronique verte et circulaire
07	Intégrité des systèmes basés sur les nouveaux paradigmes de calcul et/ou les technologies émergentes	Intégrité, test, robustesse, adaptabilité
07/08	Garantir la sécurité et la privauté des systèmes et des données	Circuits et systèmes sécurisés par conception, protections dédiées efficaces multi-niveau, approches exploitant de nouveaux paradigmes ou de nouvelles technologies
07/08	IA embarquée, continuum de calcul du <i>edge</i> au <i>cloud</i>	Approches décentralisées ou distribuées, nouveaux paradigmes, accélérateurs matériels

07/08	IA embarquée proche capteur	Interfaces avancées, compressed sensing, systèmes multi-capteurs, haptique, réalité étendue, calcul ultra-faible consommation
07/08	Calcul quantique grande échelle	Méthodes de conception et de simulation, interfaces cryogéniques, démonstrateurs matériels
07/08	IoE pour Industrie 4.0 durable, intelligente et agile	Systèmes de capteurs distribués à grande échelle, environnements industriels pilotés par les données
07/08	Convergence calcul avancé et technologies du futur	Méthodes holistiques prédictives, co-conception système-technologie émergente
08	Convergence électronique-photonique	Photonique silicium, in-network computing, communications 6G/THz
08	Systèmes de calcul à base de matériaux émergents	Approches et réalisations multi-échelle, du dispositif au système

e. Des actions phares pluridisciplinaires : IA embarquée et calcul quantique

L'importance que revêt l'IA dans le monde des systèmes embarqués fait que la dynamique industrielle est très significative sur plusieurs sujets relatifs, tels que les accélérateurs embarqués. Il est alors crucial de bien définir les sujets et les modalités de collaboration, et de mettre en place des instruments pour favoriser l'articulation formation – recherche – industrie.

Les opportunités essentielles s'entendent thématiquement, dans le défi de l'interdisciplinarité en catalysant des recherches à l'interface avec certaines communautés (d'autres GdR comme BioComp, MADICS, IA ; la société savante francophone de l'apprentissage machine ; les 3IA et notamment MIAI, le seul Institut avec une action sur l'IA embarquée), en s'appuyant sur le spectre très large des compétences en IA dans le GdR SoC². Le GdR est également au cœur des autres défis transverses liés à l'IA telle que les questions de sécurité, certifiabilité qui créent donc autant d'opportunités d'innover. Le contexte actuel du plan de relance avec l'emphase sur l'aspect IA embarqué est également largement propice à mettre en avant des projets ambitieux avec des partenaires industriels.

Sur un autre plan et à plus long terme une action pluridisciplinaire est à construire autour du calcul quantique. Mettant en jeu des mécaniques physique très poussées, des actions de structuration vers cette communauté et la communauté SoC² permettraient de commencer à combler le retard accumulé sur cette thématique en travaillant sur des approches en rupture sur ces technologies prometteuses, sur le plan des interfaces, des paradigmes et technologies de calcul, et des outils de conception.