

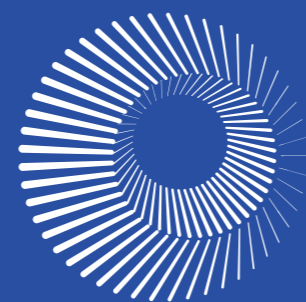
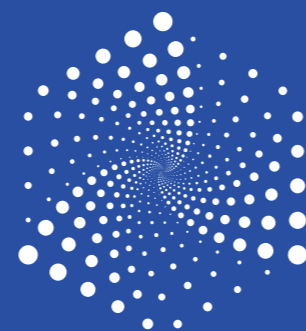
Éditorial

« Le rapport d'activité scientifique 2025 témoigne de la richesse de nos travaux scientifiques et met en lumière les résultats marquants issus de projets à la pointe de l'innovation. »

ALEXANDRE BOUNOUH
DIRECTEUR DU CEA-List



© CEA



L'ANNÉE 2025 a été pour le CEA-List une période de consolidation exceptionnelle et de projection vers l'avenir. Dans un paysage marqué par une accélération technologique sans précédent — de l'essor de l'IA générative à l'émergence de la *Physical AI* en passant par le calcul quantique — notre mission d'anticiper et d'accompagner les ruptures au service d'une innovation responsable n'a jamais été aussi pertinente.

L'excellence de notre recherche a été confirmée par le Comité de l'HCERES dont le rapport d'évaluation apporte une analyse fine et pertinente de nos activités. Il souligne notre positionnement unique, véritable pont entre la recherche amont et la compétitivité industrielle nationale. Fort de cette reconnaissance, qui place notre recherche au rang mondial avec une activité contractuelle exceptionnelle, nous avons engagé une transformation profonde de nos programmes.

Ce travail s'est cristallisé autour de l'évolution de nos programmes dans le cadre de la phase 2 de notre stratégie « Reveal ». Cette programmation ambitieuse repose sur la stratégie de Continuité Numérique, levier important de la transformation numérique de l'industrie française et européenne. Elle se traduit par le lancement de quatre grands programmes technologiques dès janvier 2026 autour des modèles de monde pour la robotique intelligente, de l'orchestration d'IA agentiques distribuées, de plateforme décentralisée d'apprentissage de spécialisation de LLM et d'instrumentation multimodale pour le contrôle. Ces programmes ont vocation à accélérer le développement de nos technologies et leur déploiement dans l'industrie en s'appuyant sur des ruptures scientifiques qui ont déjà fait l'objet de publications scientifiques de rang A et des brevets d'invention renforçant notre portefeuille de 1850 brevets actifs.

D'autres grands programmes dans le cadre de cette stratégie d'accélération préfigurent les succès de demain. Notre engagement renforcé dans les Agences de programme du CEA (ASIC) et d'INRIA (Numérique), a permis la construction de

nouvelles initiatives comme les PEPR CAMELIA sur les composants de calcul IA haute performance, PHOENIX sur la conception de composants et EDT sur les Jumeaux Numériques. C'est également le cas au niveau européen avec le pilotage de la vision européenne sur la robotique et IA, et le lancement de la plateforme de design européenne EuroCDP dont les projets en France seront portés par le DET (Design Enablement Teams) IC-DASH piloté par le CEA-List.

Nos travaux de recherche poursuivent une finalité claire, soutenir par l'innovation les acteurs industriels. Dans ce contexte, les List Tech Days 2025 ont été un événement majeur de cette année avec plus de 500 décideurs industriels qui ont répondu présents pour découvrir nos dernières avancées, confirmant notre rôle central dans l'écosystème de la recherche française et européenne, et notre capacité à transformer la recherche en innovation concrète pour l'industrie.

Cet engagement sera au cœur des enjeux de l'année 2026 avec comme point central notre réponse à l'évolution du dispositif Carnot et son ouverture massive, notamment au Pôles Universitaires d'Innovation (PUI). Si le cadre change, la mission fondamentale reste la même : stimuler l'économie par l'innovation technologique. En 2026, le dispositif doit se réinventer pour devenir plus agile, plus réactif et encore plus ancré dans les défis contemporains (transition écologique, souveraineté numérique).

Enfin, 2026 sera l'occasion de faire évoluer nos modèles de collaboration industrielle au travers de nouveaux « programmes affiliés », comme le DEFI Lab, et de nos plateformes technologiques au service de nos projets partenariaux.

Le rapport d'activité scientifique 2025 témoigne ainsi de la richesse de nos travaux scientifiques. Il met en lumière des résultats marquants issus de projets à la pointe de l'innovation et illustre notre capacité à transformer la recherche amont en solutions concrètes pour la compétitivité industrielle, avec des avancées significatives dans tous les domaines du numérique.

Sommaire

1 Éditorial

4 Introduction à l'ensemble du rapport scientifique

6 Chiffres clés 2025

7 Prix et distinctions

10 Nos outils pour soutenir la recherche

12 Les points marquants de l'année 2025

12 Une stratégie européenne

13 Les grands projets

14 Nos plateformes

15 Transferts et applications industriels



16 Jumeau numérique et ingénierie système

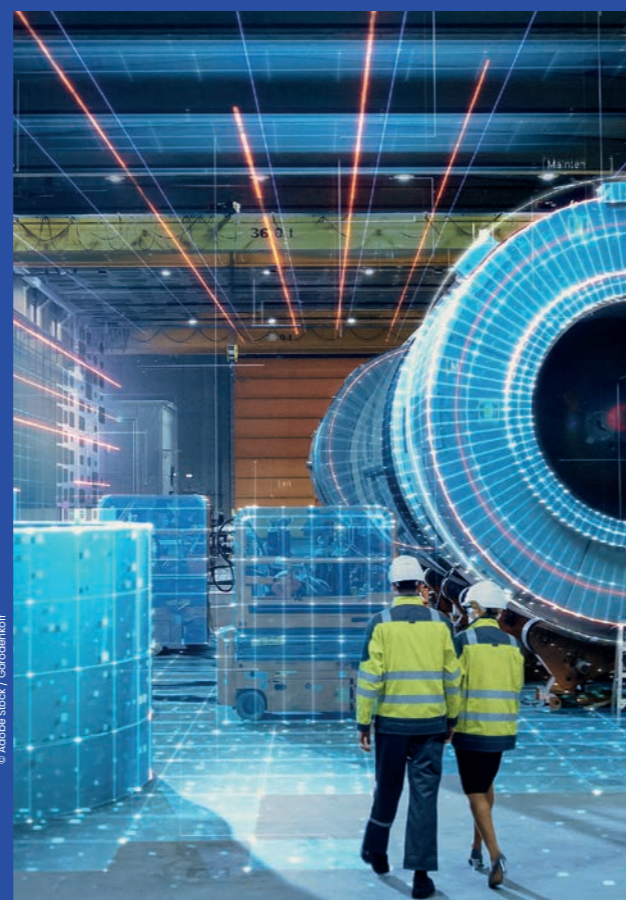
18 Le jumeau numérique au service des chaînes de productions adaptables et dynamiques

20 Accélération de la simulation multi-physique par IA : une architecture pour la mécanique des structures non-linéaires

22 Modélisation par éléments finis de la propagation ultrasonore dans les milieux visco-élastiques

24 Les jumeaux numériques, un impact tout sauf virtuel

26 Une méthodologie basée sur l'apprentissage automatique pour une modélisation rapide et efficace de la consommation de puissance des architectures numériques



© Adobe Stock / Gorodenkoff



28 Les technologies de l'intelligence artificielle

30 MiRAG : génération augmentée de récupération multi-niveau pour répondre à des questions visuelles

32 Découvrir le monde sans annotations avec xMOD : collaboration entre vision 2D et 3D

34 Analyse de scène 3D à partir de requêtes en langage naturel

36 Détection de mouvement à faible latence avec des graphes d'événements

38 Vers des systèmes d'acquisition intelligents

40 L'adaptation décentralisée de domaine multi-source dans le cadre d'une approche fédérée : pour un apprentissage collaboratif sans serveur

42 Apprentissage adaptatif pour contrer l'oubli catastrophique et la dérive de concept dans un contexte fédéré : application à une gestion optimisée des bornes de chargement électriques alliant prédiction et confidentialité

44 Analyser et réduire les biais politiques dans les grands modèles de langue

46 La vérification des réseaux de neurones : un défi à relever

48 Incertitude dans les simulations Monte-Carlo aidées par l'IA



50 Le calcul et les systèmes distribués

52 Auditer sans rien laisser fuiter

54 Simulation d'attaques sur systèmes décentralisés

56 Faire de l'intégration 3D une réalité

58 Système intégré de mémoire non-volatile embarquée avec accès robustes et optimisés

60 Cas d'usages et applications du calcul quantique

62 Qbricks : vers la vérification formelle de programmes hybrides classique/quantique



64 L'instrumentation numérique et l'usine du futur

66 Estimation de la configuration articulaire d'un préhenseur pluridigital à partir de nuages de points 3D et d'IA

68 Une interface de programmation cognitive

70 Lubrification active, frein haptique kinesthésique et vibrotactile

72 Réseaux de Bragg sur fibre optique pour la mesure ultrasonore à haute température

74 Des détecteurs quantiques lèvent le voile sur les neutrinos

76 Techniques de mesures électrochimiques pour la recherche de nitrites

78 Surveillance continue par courants de Foucault de la dégradation par fatigue dans un échantillon en acier non entaillé

80 Algorithmes d'IA pour la spectrométrie gamma dédiée à la mesure de terrain

82 Photoneutrons et photofission, armes scientifiques contre les trafics



84 Cybersécurité et numérique de confiance

86 Vérification incrémentale de logiciels, pour la cybersécurité

88 Contrôle d'attaquant et priorisation de bugs

90 Chiffrement rapide et homomorphe complet

92 Accélérateur étroitement couplé à un coeur RISC-V pour la cryptographie post-quantique

Introduction



POUR MENER ses recherches au meilleur niveau, le CEA-List s'appuie sur l'ensemble des outils de collaboration disponibles en interaction étroite avec les mondes académique, institutionnel et industriel, au niveau national et européen. Cette stratégie de collaboration est renforcée par l'investissement et le déploiement par le CEA-List de plateformes d'expérimentation et de développement avec les technologies les plus avancées, favorisant l'accélération de l'innovation industrielle.

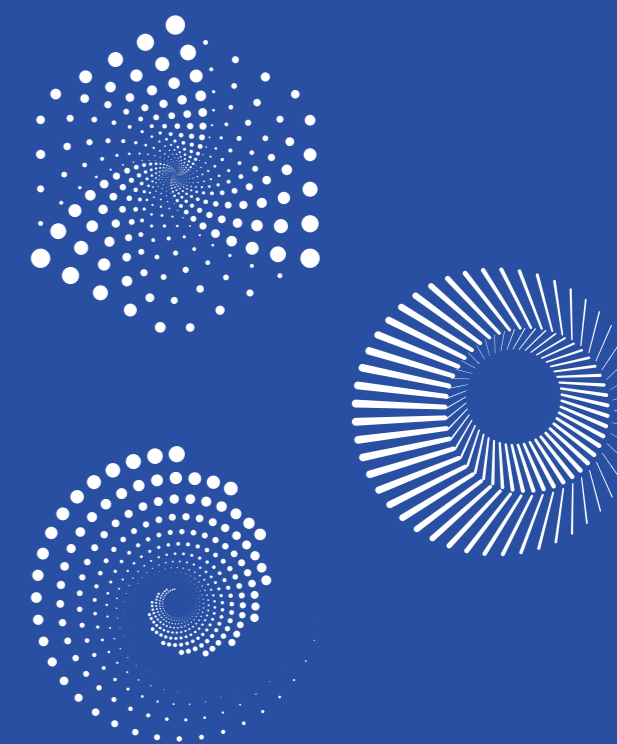
La recherche du CEA-List se développe sur un ensemble d'axes stratégiques liés à l'accélération du développement du numérique. Elle vise une maîtrise complète de l'ingénierie des systèmes et des données, de la définition des exigences à la mise au point et la qualification des systèmes. Son approche s'appuie sur l'élaboration d'outils mathématiques et logiciels de modélisation, de construction de jumeaux numériques et de simulation intégrant une démarche d'éco-conception.

Au cœur du développement de nouveaux usages en vision, traitement du langage et des signaux physiques, l'intelligence artificielle est également un axe majeur de recherche du CEA-List. Les travaux visent à construire des approches plus frugales, performantes, sûres et robustes, intégrables dans les systèmes industriels. Le déploiement durable et à grande échelle du numérique nécessite toutefois des solutions de calculs adaptées, allant des architectures et composants pour l'IA à la continuité des calculs du cloud à l'embarqué, jusqu'au quantique.

Carburant essentiel du numérique, la capture et le traitement des signaux physiques et des données mobilisent une part importante des recherches de l'Institut, en s'appuyant sur ses expertises dans les domaines du nucléaire, du contrôle qualité et de la perception, notamment. L'objectif est la construction de chaînes d'instrumentation et de perception complètes, du capteur aux systèmes industriels, équipements, robots autonomes et usines du futur.

La cybersécurité est, aujourd'hui plus que jamais, un enjeu central de sécurité et de souveraineté. Nos recherches sont ainsi centrées sur le développement de solutions, d'outils et de technologies pour la détection, l'analyse et la qualification de la sécurité, ainsi que pour le développement de solutions de protection au niveau des architectures et des technologies de calcul. Elles se déclinent sur l'ensemble des axes et domaines d'action abordés par le CEA-List.

Cette stratégie de recherche et ses modes d'action permettent aux équipes de l'Institut de démontrer son excellence en recherche et en innovation, illustrée par son rayonnement scientifique et ses transferts industriels.



© P. Charlet / CEA

© A. Aubert / CEA

Chiffres clés

1000
personnes

dont

40
HDR

160
doctorants

3
senior fellows
(intelligence artificielle,
architectures de calcul,
ingénierie système)

12
fellows

63
experts seniors

113
experts

36
directeurs
de recherche

PRODUCTION
SCIENTIFIQUE

300
publications
de rang A

1850
brevets

65
nouveaux
brevets en 2025

VALORISATION
INDUSTRIELLE

33
startups créées
depuis 2003

137
licences

200
partenaires
industriels

RESSOURCES

500
projets par an

137 M€
de budget annuel

80%
de ressources externes

Prix et distinctions 2025

Les prix et distinctions obtenus au cours de l'année témoignent de l'excellence scientifique de nos chercheurs, de leur capacité à produire des avancées majeures dans leurs domaines respectifs et à contribuer activement aux grands enjeux scientifiques, technologiques et sociétaux.

➤ **L'outil PyRAT vainqueur du « Défi sécurisation des IA »** organisé le 30 avril 2025 par l'AMIAD (Agence Ministérielle pour l'IA de Défense).



Remise des prix du défi « Sécurisation de l'IA » - © COMCYBER

➤ **L'outil Colibri vainqueur des catégories QF_FP (24s), QF_FPLRA (24s et unsat)** pour le raisonnement automatique et la preuve de programmes lors de la compétition SMT-COMP'25, 11 août 2025, Glasgow, Royaume-Uni.

➤ **1^{ère} place ex-aequo sur l'extraction d'informations few-shot, et 2nde place sur le fine-tuning, pour les données fermées,** au challenge EvalLLM, organisé par l'AMIAD - 30 juin 2025, Marseille, France.



Gant haptique Wavy

➤ **Prix de la meilleure démonstration lors de la conférence I3E VR 2025,** pour le gant haptique Wavy, Mars 2025, Saint-Malo, France.

➤ **Premier prix du Nuclear Innovation Prize** décerné par EURATOM au CEA-List et l'ASNR dans la catégorie « Reactor Safety System » pour la proposition COMOS lors de la 11^{ème} Conférence FISA-EURADWASTE, Mai 2025 qui récompense la contribution scientifique du CEA-List en spectrométrie Raman par fibres optiques, appliquée à la gestion du risque hydrogène en situation d'accident grave.



Premier prix du Nuclear Innovation Prize

➤ **Lauréat des concours d'innovation de l'Etat 2024-2025** soutenus par France 2030, dans la catégorie i-Nov pour Snowpack, startup du CEA-List.



➤ **2nde place au défi TextMine'25,** défi proposé par Airbus Defense & Space, visant l'extraction de relations entre entités à l'échelle du document, pour des documents en français simulant des rapports de renseignement, lors de la conférence EGC 2025, 28 janvier 2025, Strasbourg, France.



Prix et distinctions 2025

- > **2nd place générale au CheckThat! 2025** Subjectivity challenge, pour 'Evaluating LLMs as Detectors of Bias and Opinion in Text'.
1^{ère} place pour les langues Arabe et polonais lors de la conférence CLEF2025 – Septembre 2025, Madrid, Espagne.
- > **2^e prix d'innovation de transfert technologique à ERF2025** pour le robot d'inspection des réseaux d'assainissement PREDIRE, Mars 2025, Stuttgart, Allemagne.
- > **3^e place au Perception Bin Picking Challenge**, catégorie « one-shot method » à la conférence I3E CVPR, Juin 2025, Nashville, États-Unis.
- > **3^e place au challenge Welding Quality Detection** lors du Trustworthy AI Summit 2025 – 16 septembre 2025, EDF Lab, Saclay, France. Défi axé sur la classification automatisée d'images de soudures prises sur les lignes de production automobile (cas d'usage industriel Renault).
- > **Obtention de la chaire d'excellence INSITU** du Cluster IA MIAI de l'UGA portée par Marielle Malfante. Elle doit contribuer à l'IA embarquée et aux capteurs intelligents pour la surveillance de l'environnement.
- > **Best Artifact Award à Dimitri Kokkonis** et Michaël Marcozzi et leurs co-auteurs Emilien Decoux et Stefano Zacchiroli, pour l'outil ROSA, qui permet de débuser des failles logicielles de type backdoor à l'aide de fuzzing lors de la conférence ICSE'25, Mai 2025, Ottawa, Canada.
- > **Notable Review Award à Sébastien Bardin** lors de la 34^{ème} conférence Usenix sur la sécurité et la confidentialité des systèmes et réseaux informatiques, Juillet 2025, Seattle, États-Unis.
- > **Distinguished Review Award à Sébastien Bardin** lors de la conférence NDSS, Février 2025, San Diego, États-Unis.
- > **Best Paper Award à James Miranda et Ansgar Radermacher** et leurs co-auteurs Lukasz Mazur, Nenad Petrovic, Robert Rasche et Alois Knoll, pour l'article "Querying Large Automotive Software Models: Agentic vs. Direct LLM Approaches" sur l'application de l'IA pour gérer les modèles automobiles complexes, lors de la conférence GACLM 2025, Août 2025, Valence, Espagne.



- > **Best Student Paper Award à Maxence Perion** pour son article « Des Quorums Furtifs pour Contrer l'Adversaire Adaptatif et Dépenser Partiellement des Jetons » à la conférence internationale en systèmes distribués Algotel, Juin 2025, Saint Valéry-sur-Somme, France.
- > **Best Student Paper Award à Marion Pontreau** à I3E WHC 2025 (World Haptics) – Juillet 2025, Suwon, Corée du Sud.
- > **Best Paper Award à Mohamed Amine Zhiri**, doctorant pour l'article « HT-NoC : Reconfigurable High Throughput Network-on-Chip for AI Dataflow Accelerators » lors de conférence ARC 2025, Avril 2025, Séville, Espagne.
- > **Prix de la meilleure présentation de vulgarisation "Speed-poster" attribué à Matthieu Lemerre** au Workshop Interdisciplinaire pour la Sécurité Globale (WISG) de l'ANR, Mars 2025, Plateau de Saclay, France.
- > **Prix du meilleur poster à Ablaihan Uteпов**, lors de l'ICRM 2025, Mai 2025, Paris, France.
- > **Best Paper Award à Eduardo Tomasi Ribeiro**, doctorant pour son article « HPC Workload Analysis Using Distributed Cross-ISA Binary Instrumentation » lors de la conférence DCIS 2025, Novembre 2025, Santander, Espagne.
- > **Outstanding Paper Award à Lilia Rouizi**, doctorante, pour son article « Revisiting Timing Anomalies in Predictable In-Order Pipelines » lors de la conférence ECRTS 2025, Juillet 2025, Lund, Suède.
- > **Prix « Doctorant STIC du Plateau de Saclay » à Dorian Lesbre** pour son article « Compiling with Abstract Interpretation » sur la validation formelle de logiciels.
- > **Prix de la meilleure présentation à Aïcha Boukhari** pour la validation formelle de logiciels lors de la conférence JITA'25, Octobre 2025, Plateau de Saclay, France.
- > **Prix du meilleur poster attribué à Sasha Cuau** sur la robustesse des IA à JITA'25, 2nd, 16 octobre 2025, Plateau de Saclay, France.



Nos outils pour soutenir la recherche

POUR SOUTENIR les équipes dans leurs activités de recherche, il est essentiel de favoriser leur implication dans la communauté académique et leur investissement dans des actions de recherche exploratoires et fondamentales. Les programmes de recherche prioritaires (PEPR), mis en place dans le cadre de la stratégie France 2030, se positionnent ainsi sur des thèmes de recherche à moyen et long terme favorisant l'émergence de ruptures scientifiques. Leur caractéristique unique est d'offrir aux équipes les moyens de mener des projets ambitieux et de structurer, à l'échelle nationale, des communautés sur des thématiques stratégiques. Le CEA-List, fortement mobilisé sur ces initiatives, contribue aujourd'hui à 19 PEPR, soit en participant ou coordonnant des projets, soit en pilotant des PEPR. Ainsi, le CEA-List dirige avec d'autres grands organismes de recherche (CNRS, Inria, IMT, notamment) deux programmes stratégiques pour ses activités : le PEPR Intelligence Artificielle et le PEPR Cloud. Parallèlement, il pilote et contribue à plusieurs projets clés des PEPR Cybersécurité et Réseaux du futur. Au-delà du développement de collaborations au sein des projets, cette transversalité entre les organismes de recherche favorise l'échange de points de vue et l'émergence de synergies sur les sujets communs ou connexes à plusieurs PEPR.



LES AGENCES DE PROGRAMMES

La mise en place des agences de programmes nationales complète le rôle des PEPR en déléguant à des organismes scientifiques de référence la mise en place de programmes de recherche et d'innovation répondant à des enjeux économiques clés. Le CEA-List est tout particulièrement impliqué dans deux agences : l'agence Architecture, Systèmes, Infrastructures et Composants portée par le CEA et l'agence Numérique (usages, logiciels et algorithmes), portée par l'Inria.

Celles-ci ont lancé de nouveaux programmes structurants en 2025 pour relever les défis de compétitivité industrielle et de souveraineté :

- Les composants haute performance et basse consommation pour l'IA (CAMELIA). Co-porté par le CEA-List avec l'Inria, ce programme exploitera les avancées du PEPR IA sur la frugalité et l'IA embarquée.
- Les systèmes souverains de calcul hétérogènes, haute performance (PHOENIX). Co-porté par le CEA-List et le CNRS, il s'appuie sur les programmes en microélectronique associant technologies, architectures et environnements de conception.
- L'ingénierie système avancée pour les jumeaux numériques (EDT - Engineering Digital Twins) pour la constitution d'une plateforme d'ingénierie systèmes.
- L'évaluation et la sécurité de l'IA, où le CEA-List contribue sur la cybersécurité de l'IA, la sûreté des systèmes à base d'IA, et la maîtrise des biais en IA générative.

PEPR IA : PROGRAMME DE RECHERCHE PRIORITAIRE SUR LES FONDEMENTS DE L'INTELLIGENCE ARTIFICIELLE



73 M€ - 9 projets ciblés dont 5 copilotés par le CEA-List

200 chercheurs et 200 recrutements de thésards et postdocs sur tout le territoire national

Thématiques prioritaires
Frugalité et sûreté de l'IA, IA embarquée et décentralisée

<https://www.pepr-ia.fr/le-pepr/>



PEPR CLOUD : PROGRAMME DE RECHERCHE PRIORITAIRE SUR LES FONDEMENTS DU CLOUD



51 M€ - 7 projets ciblés dont 2 copilotés par le CEA-List

192 chercheurs et 250 recrutements de thésards et postdocs sur tout le territoire national

Thématiques prioritaires
Continuité numérique, sécurité, virtualisation, orchestration

<https://pepr-cloud.fr/fr/>



Une stratégie européenne

Pour renforcer son écosystème scientifique à l'échelle européenne, le CEA-List met en œuvre une stratégie ambitieuse de collaboration et de positionnement.

Acteur dynamique de la recherche et de l'innovation, il contribue à la représentation du CEA dans plusieurs associations et partenariats clefs pour nos domaines d'expertise : EFFRA pour l'Usine du Futur, ADRA pour l'IA, Data et robotique, BDVA sur le Big Data, ECSO sur la Cybersécurité, INSIDE pour les systèmes numériques intelligents et Eclipse pour le logiciel open source. Nous sommes également engagés dans cinq réseaux d'excellence, sur l'IA pour le secteur des médias (AI4Media), l'IA de confiance

(TAILOR), les systèmes de calcul (HiPeac), l'IA reproductible (euRobin) et l'IA distribuée de confiance (dAIEDGE).

Chaque année, le CEA-List participe à ou coordonne environ 100 propositions pour répondre aux différents appels à projets, au sein des programmes Horizon Europe, Digital Europe, le Fonds européen de défense, Euratom, EuroHPC ou d'autres Joint Undertakings (comme ChipsJU). Les 94 projets européens actuellement menés couvrent des secteurs divers, de l'industrie au numérique et à la santé, et de la sécurité civile à la défense, l'agriculture et l'environnement. En cohérence avec notre programme d'éco-innovation, nous mettons nos expertises au service des défis sociétaux

de l'économie circulaire, via les projets de passeport numérique des produits, ou le partage sécurisé de données industrielles au niveau européen.

L'Europe a besoin d'installations de référence de classe mondiale et à fort impact pour permettre aux entreprises industrielles, de toute taille, de tester des concepts et produits innovants basés sur l'IA dans des conditions représentatives de la production. Le projet AI-MATTERS (*Test & Experiment Facilities*) a ainsi permis au CEA-List de créer PRISM, une infrastructure unique de test et d'expérimentation. Au travers des services fournis aux PME, PRISM contribue à renforcer la compétitivité de l'industrie manufacturière européenne.

L'éco-innovation

Attaché aux valeurs d'éco-responsabilité, le CEA-List a initié depuis plusieurs années une **démarche d'éco-innovation**, aujourd'hui incarnée dans un programme de recherche dédié. Cette démarche est décrite dans un article¹ qui définit les notions de **SustainableTech**⁽¹⁾ (visant à réduire l'impact environnemental et sociétal direct de la technologie) et de **Tech-4Sustainability** (utilisant la technologie pour contribuer à résoudre des problématiques environnementales et sociétales et évaluer les impacts indirects), ainsi que les méthodologies sous-jacentes reposant sur des normes existantes. La démarche d'éco-innovation implique d'explorer les deux volets à la fois : il s'agit d'être en mesure de proposer une technologie et d'évaluer son bénéfice net, en tenant compte aussi bien de ses impacts que de ses effets positifs sur l'environnement.

Les équipes du CEA-List généralisent progressivement cette démarche, notamment en l'appliquant à la mise en place du Passeport Produit Numérique (DPP) et à l'outillage logiciel de la démarche d'évaluation des impacts environnementaux (ACV). Un autre exemple est l'application de technologies du CEA-List pour le recyclage de matériaux critiques et l'identification d'indésirables dans les flux de déchets, dans le contexte de la gestion des DEEE (Déchets d'Équipements Électriques et Electroniques).

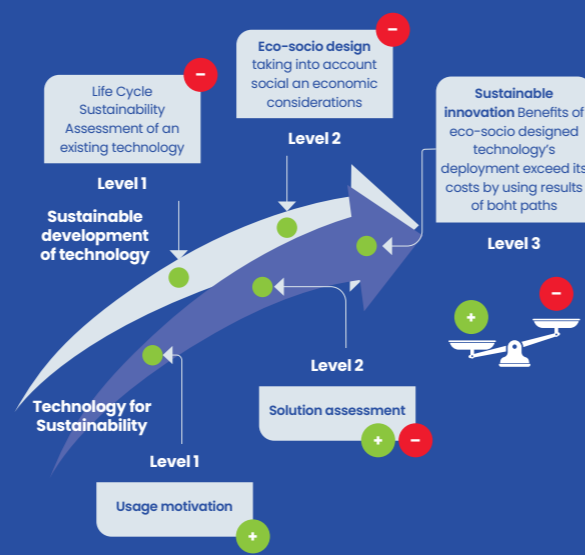


Figure 1 : Illustration de la démarche d'éco-innovation.

¹ Sustainable technology and Technology for sustainability: The paths towards Eco-innovation, Chiara Sandionigi, Jean-François Berrée, Maxime Peralta, Ariane Piel, Bénédicte Robin, Electronics Goes Green 2024.

Les grands projets

Les compétences scientifiques de niveau international et la culture industrielle des équipes du CEA-List lui ont permis de construire de grands projets visant la capitalisation et le déploiement de ses innovations.

Citons en particulier :

- **Le projet européen STARLIGHT (2021-2025)**, qui a permis de structurer une communauté européenne (50 partenaires, 18 pays) durable et multidisciplinaire autour de l'IA pour la sécurité, réunissant forces de l'ordre, chercheurs, industriels et experts éthiques. Grâce à une méthodologie de co-innovation (CODEV) et huit cas pilotes, 50 outils d'IA interopérables ont été développés pour répondre aux besoins d'anticipation des menaces, d'investigation et de cybersécurité. Le projet a établi un cadre robuste de conformité éthique, juridique et réglementaire, incluant la protection des données, la traçabilité et l'alignement avec l'AI Act.
- **DeepGreen(deepgreen.ai)** : le projet vise la construction d'une plateforme souveraine ouverte pour le déploiement de l'IA en embarqué. S'appuyant sur un consortium de plus de 20 partenaires, ses avancées sont intégrées dans la plateforme open source Aidge hébergée par la fondation Eclipse (eclipse.dev/aidge/). Destiné à favoriser le déploiement de l'IA sur une grande variété de cibles du marché ou émergentes, le projet contribue à redonner aux industriels une autonomie dans le choix des technologies qu'ils souhaitent utiliser.



© P. Chaillet / CEA

- **Un ensemble de projets coordonnés pour concevoir des systèmes de calculs haute performance** du cloud à l'embarqué, comme les projets EPI et MOSAICS-LP pour des architectures de calculs hétérogènes à base de chiplets adaptables selon les besoins d'usage, ou le projet européen CHASIS pour développer les architectures et outils habilitants le Software Defined Vehicle.
- **Le programme Q-Loop de l'IRT Nanoelec** relève le défi du passage à l'échelle de l'électronique de contrôle de l'ordinateur quantique avec correction d'erreur. Il a permis l'établissement d'une approche outillée de modélisation système (MBSE basé SysML V2) d'ordinateur quantique et sa démonstration complète sur une ligne d'atténuation. Les travaux se poursuivent pour enrichir les modèles sur l'ensemble de la chaîne et intégrer visualisation et vérification de propriétés.

- **UDD@Orano** : avec un consortium de 12 partenaires, ce projet collaboratif vise à intégrer les technologies numériques avancées pour accélérer la modernisation des usines du cycle du combustible nucléaire. Les expertises du CEA-List ont contribué au développement de onze prototypes, matériels et logiciels répondant à différents besoins critiques : localisation de sources radioactives, comptage neutronique, dosimétrie gamma aux extrémités, détection d'actinides en solution, analyse de spectres gamma complexes et recalage de modèles de simulation.

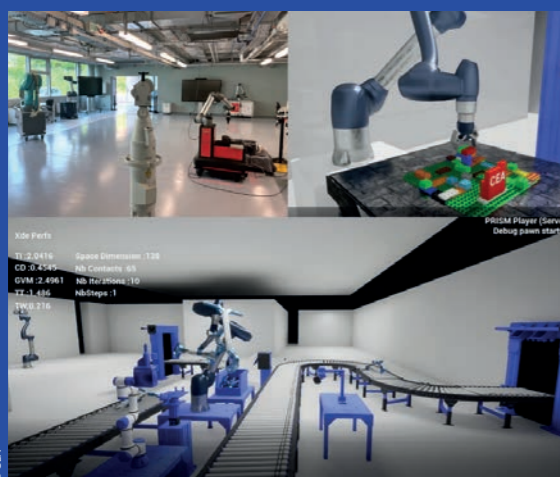
Nos plateformes



© Y. Audric / CEA

Pour soutenir ses recherches et faciliter le transfert de ses technologies vers l'industrie, le CEA-List a mis en place des plateformes d'innovation uniques.

Elles offrent un cadre structuré, où ingénieurs-chercheurs et industriels peuvent collaborer efficacement pour développer des solutions innovantes basées sur l'état de l'art du numérique. En mutualisant les compétences, les outils et les ressources technologiques, ces plateformes permettent d'accélérer la montée en maturité des solutions et de préparer les conditions d'un déploiement industriel.



© CEA

L'infrastructure PRISM rassemble ainsi des moyens numériques uniques pour tester et expérimenter en conditions quasi-réelles des solutions technologiques pour l'industrie du futur. Les projets de recherche qui y sont menés couvrent les domaines de la robotique/cobotique, du jumeau numérique, de l'IA et de la donnée, de l'ingénierie logicielle, de la continuité numérique et de la fabrication additive. Les technologies ayant atteint un niveau de maturité suffisant sont intégrées dans des « capacités » opérationnelles au sein de la plateforme, facilitant leur expérimentation sur des cas d'usage industriels.

En combinant et capitalisant ses briques technologiques au sein de ses plateformes d'innovation, le CEA-List contribue à une recherche plus agile et plus impactante, ainsi qu'à l'émergence plus rapide de solutions concrètes, durables et compétitives pour relever les défis industriels de demain.

Transferts et applications industriels

Pour accomplir sa mission de soutien à la compétitivité des entreprises, le CEA-List se doit de cultiver excellence scientifique et culture industrielle, pour transformer l'innovation technologique en solutions à haute valeur ajoutée. Les grands projets de recherche nationaux et européens et le développement de plateformes technologiques de haut niveau s'inscrivent dans cette stratégie, avec des résultats tangibles de valorisation des innovations de l'Institut. Voici quelques avancées marquantes en 2025.



© Cyril Cortez

SADE et le CEA ont dévoilé sur le salon Pollutec le robot TUBOCONTACT SD.

1. SADE : première mondiale en robotique

Le CEA-List et SADE ont développé TUBOCONTACT SD, un robot autonome pour réhabiliter les réseaux d'eau potable. Présenté à Pollutec en octobre 2025, il inspecte et réhabilite les conduites en fonte via des chantiers de courte durée, peu invasifs, sans tranchée et à coûts maîtrisés, et peu impactants en termes de nuisances.

2. SNCF Réseau et Compagnie des Signaux : homologation de systèmes logiciels pour concevoir les postes d'aiguillage du futur

Depuis 2021, le CEA-List collabore avec la Compagnie des Signaux pour développer l'Atelier Métier de Signalisation (AMS). En 2025, l'AMS et son module de génération de code ont été homologués pour les postes d'aiguillage ferroviaire informatiques.

3. TDK : mémoires spintroniques, une technologie clé de mémoire pour les systèmes d'IA

TDK et le CEA-List ont renouvelé leur collaboration pour explorer de nouvelles architectures de calcul neuromorphique basées sur les spin memristors. Cette technologie permettra aux futurs systèmes d'IA d'apprendre et de s'adapter plus efficacement, avec une consommation d'énergie divisée par cent relativement aux autres composants.

4. MBDA : premier prix d'innovation

Depuis 2020, le CEA et MBDA collaborent sur l'IA embarquée. En 2025, leurs travaux communs remportent le premier prix des Innovation Awards MBDA.

Ces collaborations fructueuses traduisent l'impact concret et positif de la recherche dans l'industrie, dans un contexte en mutation rapide. En continuant à innover et à mutualiser les efforts de R&D académiques et industriels, le CEA-List contribue à façonner l'industrie de demain.

01

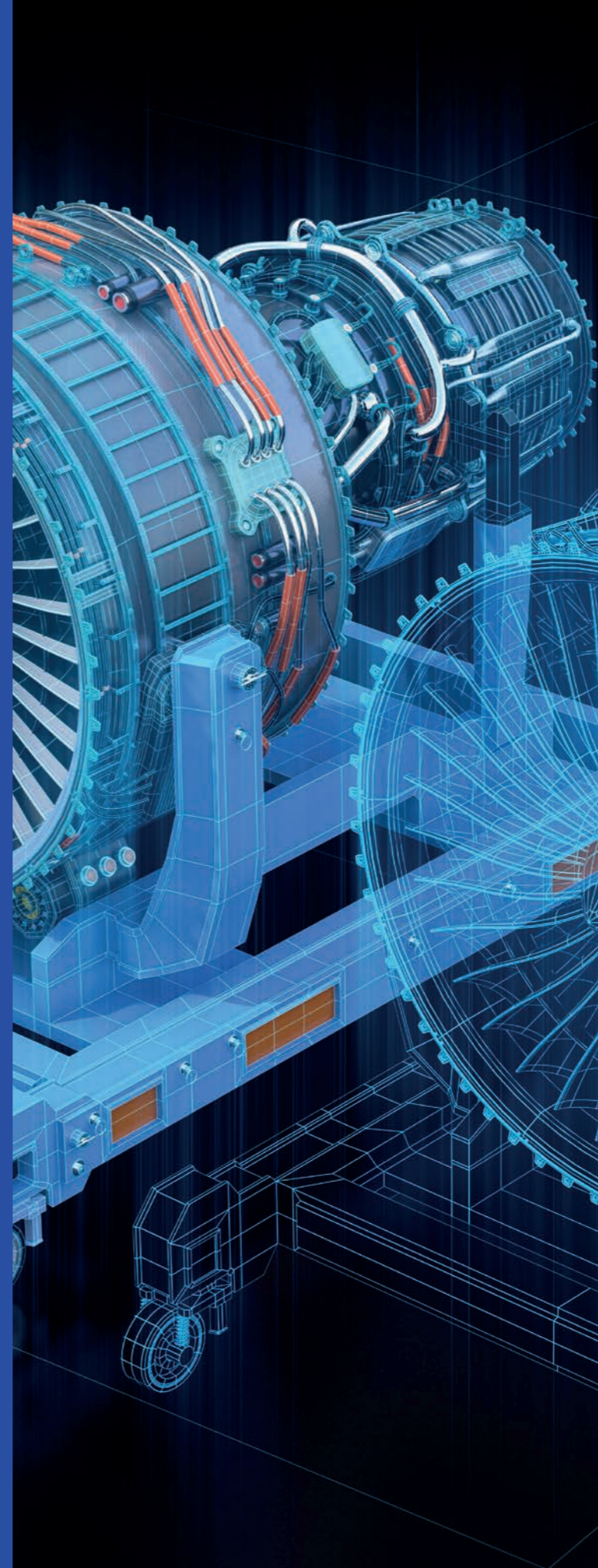
Jumeau numérique et ingénierie système

Brique essentielle de l'ingénierie système, les jumeaux numériques soutiennent la transformation des systèmes industriels en offrant des outils d'aide à la décision et la continuité entre le monde numérique et le monde physique. Fondés sur la modélisation et l'exploitation des données opérationnelles, ils produisent une réplique fidèle permettant d'anticiper les aléas, et de préparer les décisions sans perturber les systèmes réels.

Le jumeau numérique est capable d'orchestrer la co-simulation multi-niveaux et multi-métiers. Il supporte une ingénierie continue, distribuée et outillée, favorisant des systèmes plus robustes, plus flexibles et plus sûrs. À terme, une interaction dynamique avec les systèmes du monde réel ouvrira la voie à des systèmes capables de s'auto-évaluer et d'assister la prise de décision en temps réel, dans un cadre maîtrisé et vérifiable. Malgré ce potentiel, son adoption reste freinée par des verrous technologiques, au cœur des travaux du CEA-List.

Il s'agit notamment d'assurer l'interopérabilité, essentielle pour permettre au jumeau numérique de s'interfacer facilement avec les systèmes existants sans charge additionnelle. Le CEA-List répond à ce défi par la conception d'une nouvelle architecture logicielle favorisant flexibilité et automatisation. Un autre défi est lié au coût élevé de l'ingénierie logicielle, qui requiert également des compétences hautement qualifiées. Les approches basées sur l'IA, développées au CEA-List, offrent un levier pour accélérer les phases de modélisation physique et de simulation tout en réduisant le niveau de technicité requis. Enfin, le CEA-List développe une méthodologie qualitative pour répondre à une question centrale liée à l'évaluation des coûts environnementaux et économiques dans le cadre de leur adoption massive dans l'industrie.

Les perspectives concernent la sécurité des communications IT/OT et les vulnérabilités inhérentes au fort lien cyber physique, l'intégration de l'IA agentique pour des jumeaux numériques distribués, autonomes et communicants, ainsi que l'évaluation de ces concepts au sein d'une continuité numérique multi-métiers fondée sur des espaces de données partagés.



16 Technologies

- 18 Le jumeau numérique au service des chaînes de productions adaptables et dynamiques
- 20 Accélération de la simulation multi-physique par IA : une architecture pour la mécanique des structures non-linéaires
- 22 Modélisation par éléments finis de la propagation ultrasonore dans les milieux visco-élastiques

24 Méthodologies

- 24 Les jumeaux numériques, un impact tout sauf virtuel
- 26 Une méthodologie basée sur l'apprentissage automatique pour une modélisation rapide et efficace de la consommation de puissance des architectures numériques

Jumeau numérique et ingénierie système

Technologies

Le jumeau numérique au service des chaînes de production adaptables et dynamiques

Un défi de l'industrie est la production à la demande, qui lui permettrait de diminuer ses coûts et son empreinte environnementale. À l'aide de sa plateforme de modélisation Papyrus for Manufacturing, le CEA-List a conçu une nouvelle architecture logicielle pour relever ce défi de flexibilité et d'automatisation. Une expérimentation avec son partenaire APS a confirmé la pertinence de la solution.

Aujourd'hui, alors que nos regards se tournent vers l'Industrie 5.0, centrée sur l'humain, la durabilité et la résilience dans la production, de nombreux défis de l'Industrie 4.0 restent non résolus, comme la production pilotée par les commandes. Celle-ci exige des chaînes de production un haut niveau de flexibilité et d'automatisation afin de satisfaire des commandes de fabrication variées, que ce soit en quantité ou en type de produit.

Pour mettre en œuvre cette stratégie, le système de production doit être capable de : (1) suivre la disponibilité des équipements industriels, (2) assigner dynamiquement des opérations de production aux équipements et (3) permettre aux nouveaux équipements d'intégrer automatiquement le processus de production en cours.

Les standards et concepts existants répondent partiellement à ces trois exigences :

- **l'AAS (Asset Administration Shell)** est un standard pour créer des jumeaux numériques d'équipements industriels avec une interface pour faciliter leur gestion et leur communication ;
- **CSS (Capability Skill Services)** est un concept qui représente les équipements (robot, tapis roulant industriel, ...) par leurs capacités - cela permet de sélectionner dynamiquement ceux qui sont les plus adaptés à une mission donnée ;



© CEA
QUANG-DUY NGUYEN
Ingénieur-chercheur



© APS
MAXIME MORIN
Responsable R&D, APS

- **PnP (Plug-and-Produce)** désigne la capacité d'un système industriel à être reconfiguré, en permettant à un nouvel équipement de rejoindre un processus de production et de collaborer automatiquement et dynamiquement avec les autres équipements.

Cependant, certaines limites de l'AAS l'empêchent d'intégrer pleinement les concepts CSS et PnP, comme l'absence de détection de nouveaux équipements ou le manque de description détaillée des capacités d'un équipement...

Pour permettre la réalisation d'une chaîne de production adaptable et dynamique, le CEA-List a développé l'architecture logicielle XAAS (AAS Extensible) en utilisant Papyrus for Manufacturing, son outil de modélisation et de génération de jumeaux numériques au standard AAS (voir figure 1).

XAAS ajoute à l'AAS deux contributions majeures :

- **des paramètres** pour la description des capacités d'un équipement,
- **quatre nouveaux services** : détection d'équipement (*Discovery*), association capacités/mission (*Matchmaking*), *configuration* de l'équipement et *Intégration* au processus en cours.

Ces contributions permettent de réaliser pleinement le Plug-and-Produce.

Le CEA-List a implémenté l'architecture XAAS sur un cas d'usage joué pour le remplissage de boîtes d'emballage. L'expérimentation a permis de valider le concept XAAS : les nouveaux équipements ont intégré automatiquement le processus de production en cours dès leur allumage et le système a choisi les robots adaptés à la réalisation d'une mission donnée.

Grâce à XAAS, il est ainsi possible de tester plusieurs configurations d'équipements pour remplir une mission et de choisir la meilleure selon des critères de sélection (temps de production, consommation d'énergie, ...).

Figure 1 : Architecture logicielle de la solution XAAS du CEA-List.

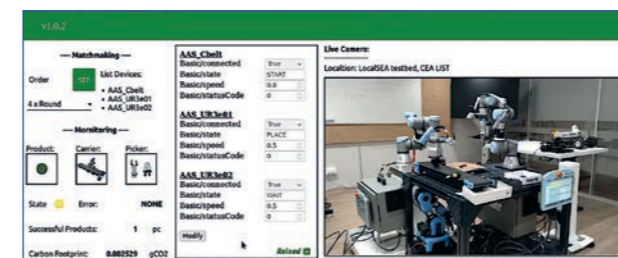
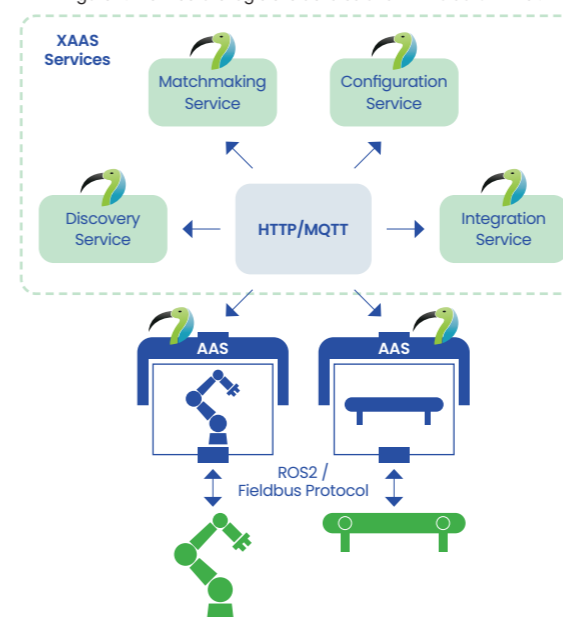


Figure 2 : Le démonstrateur.



« Notre plateforme de modélisation Papyrus for Manufacturing permet de répondre aux enjeux de flexibilité et d'automatisation de l'industrie du futur. »

••• QUANG-DUY NGUYEN

« Le projet INTEROP a confirmé le potentiel de notre architecture sur un cas d'usage représentatif. Sur cette base, APS a renouvelé son partenariat avec le CEA-List. »

••• MAXIME MORIN

- **Prototype**
« Plug&Produce dans une chaîne de production » - TRL 5.

- **Projet industriel** : APS-INTEROP, objectif : brique logicielle de « plug-and-play » dans une chaîne d'assemblage.

- **Partenaire**
APS, basé à Béziers, développe une technologie qui permet une interopérabilité « plug-and-play » pour tous les composants des systèmes de production.

- **Publication majeure**
« An AAS-based Architecture for Plug and Produce and Order-Driven Production ». Nguyen, Q.-D., Morin, M., Dhoub, S., & Navarro, N. (2025). *IFAC-PapersOnLine*, 59(10), 927-932. <https://doi.org/10.1016/j.ifacol.2025.09.157>



Jumeau numérique et ingénierie système

Technologies

Accélération de la simulation multi-physique par IA : une architecture pour la mécanique des structures non-linéaires



LEBNN (Latent-Energy-Based Neural Network) est une architecture de réseaux de neurones adaptée à la mécanique des structures hyperélastiques en statique. Elle se base sur l'apprentissage non supervisé d'une énergie de déformation « latente » de la structure, qui garantit la compatibilité de LEBNN avec des schémas de couplages et permet ainsi de l'intégrer au sein de simulations multi-physiques avec certains composants simulés par éléments finis et d'autres par des réseaux de neurones.

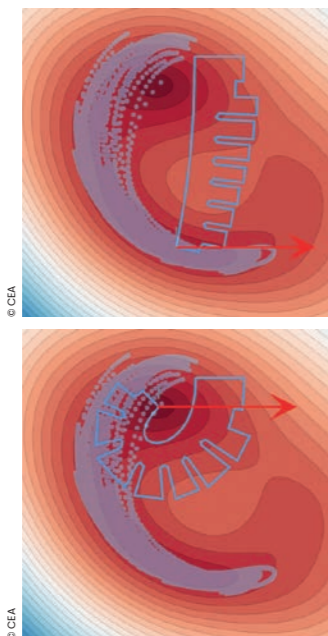


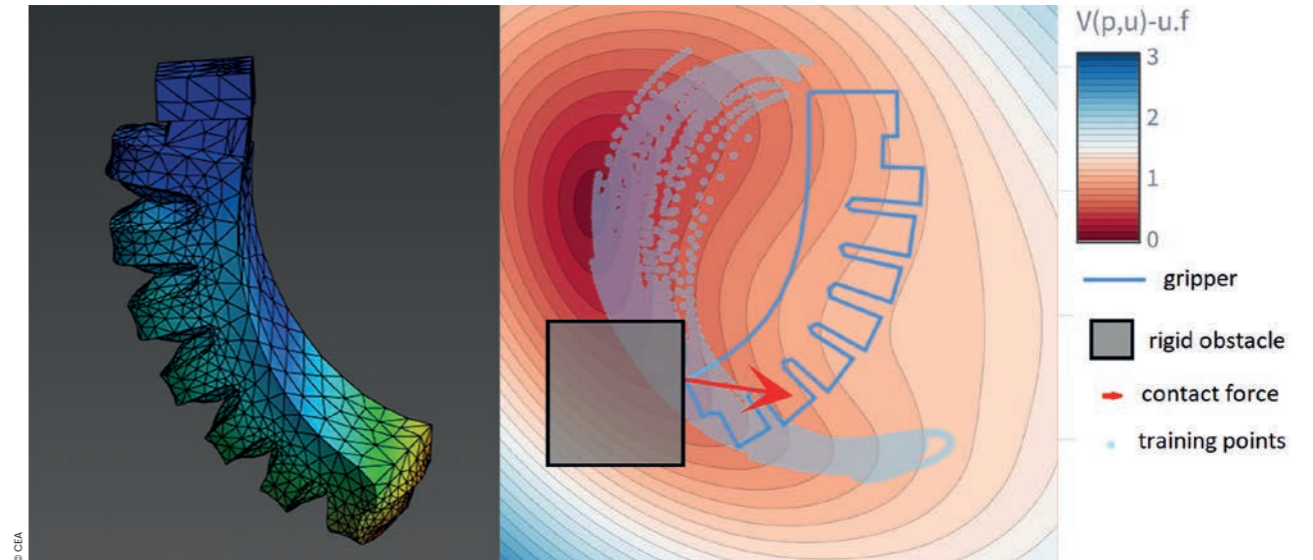
Figure 1 : Deux champs de déplacements u équilibrés par le même effort f .

En statique des structures non-linéaires, nous cherchons à trouver le champ de déplacement u d'une structure soumise à un champ d'efforts donné f , ce qui se fait classiquement par la méthode des éléments finis. Lorsque le nombre de degrés de liberté est grand, la résolution par éléments finis est trop coûteuse pour être faite en temps réel.

Pour accélérer la simulation, nous pouvons constituer une base de données de couples (u, f) et entraîner un réseau de neurones à prédire u connaissant f dans le but de le substituer au modèle éléments finis.

L'approche classique dans la littérature $u=NN(f)$ souffre de deux défauts majeurs :

- elle est incompatible avec l'apprentissage du comportement de structures fortement non-linéaires à l'énergie de déformation non-convexe (exemple Figure 1).
- elle ne permet pas de construire une matrice de raideur cohérente, donc ne peut pas être insérée dans une simulation couplée.

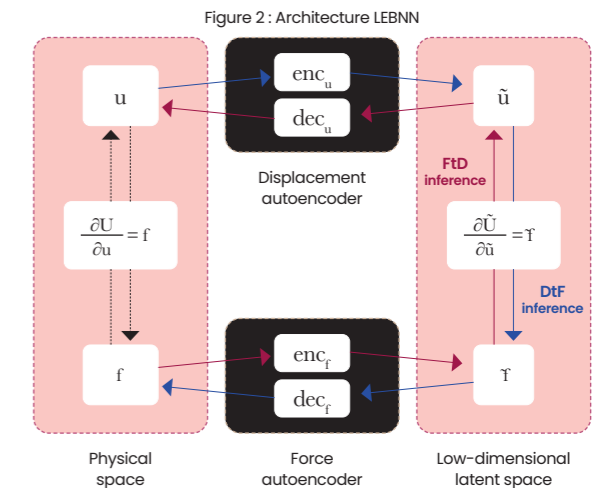


Modèle physique du préhenseur pneumatique du projet ARISE (gauche) ayant servi à entraîner un LEBNN utilisé dans une simulation du contact en temps réel (droite). Les lignes de niveaux représentent l'énergie latente.

L'architecture LEBNN résout ces deux problèmes en se basant sur l'apprentissage non-supervisé d'une énergie de déformation « latente » dans un espace de dimension faible. L'énergie de déformation latente encode les propriétés de la structure. En particulier, elle a la même topologie que l'énergie de déformation réelle de la structure (qui, elle, est généralement inconnue). La matrice de raideur de la structure peut être calculée à partir de la Hessienne de l'énergie latente. Ainsi une fois entraîné, LEBNN peut être inséré dans une simulation multi-physique et se substituer entièrement au modèle éléments finis de la structure apprise.

Dans le cadre du projet européen ARISE, nous devons réaliser une simulation interactive intégrant un préhenseur flexible en matériaux hyperélastiques dont les simulations éléments finis prennent plusieurs minutes, très loin du temps réel. Nous avons pu remplacer le modèle physique du préhenseur par un LEBNN entraîné sur une centaine de simulations, dont le temps d'inférence inférieur au dixième de seconde rend possible le couplage à un algorithme de contact tournant en temps réel.

Ainsi, LEBNN ouvre la perspective de simuler en temps réel une nouvelle classe de systèmes.



- **Cas d'usage**
Simulation temps réel de structures hyperélastiques.

- **Projet majeur : ARISE (EU).**

- **Publication majeure**

« Latent-Energy-Based NNs: An interpretable Neural Network architecture for model-order reduction of nonlinear statics in solid mechanics », L. Pottier, A. Thorin, F. Chinesta, *Journal of the Mechanics and Physics of Solids*, Vol. 194, Jan. 2025, Article 105953, pp. 1-13, <https://doi.org/10.1016/j.jmps.2024.105953>

Jumeau numérique et ingénierie système

Technologies

Modélisation par éléments finis de la propagation ultrasonore dans les milieux visco-élastiques

Pour la simulation de la propagation d'ondes visco-élastiques, les schémas explicites classiques – comme ceux utilisés dans la plateforme CIVA pour la simulation des expériences de contrôle non destructif – deviennent impraticables dès qu'apparaissent des contrastes localisés ou des maillages fins, émanant de géométries complexes. Ce travail propose deux stratégies numériques permettant de lever cette contrainte tout en garantissant stabilité et efficacité de calcul.

Pour la simulation de la propagation d'ondes visco-élastiques appliquées au contrôle non destructif (CND), la présence d'inclusions à très forts contrastes, de couches minces, ou d'éléments de maillage fortement déformés génère des difficultés dans les algorithmes standards de discrétisation en temps. Celles-ci imposent à ces algorithmes une contrainte très forte sur le pas de temps maximal autorisé assurant un calcul numérique correct. Cette contrainte est notamment liée au rayon spectral d'opérateurs de rigidité définis dans la zone « perturbée », et conduit à des pas de temps extrêmement faibles et à un coût de calcul disproportionné.

Le travail présenté ici propose, et analyse, deux familles de stratégies numériques permettant de découpler la stabilité globale de l'algorithme de discrétisation en temps de la raideur locale, dans la zone « perturbée ».

Premièrement, nous étendons aux modèles visco-élastiques les schémas localement implicites (LI), formulés en appliquant un schéma implicite uniquement pour les opérateurs locaux. Nous sommes capables de prouver la stabilité de l'algorithme final, par une méthode d'énergie, indépendamment des opérateurs restreints à la zone « perturbée ». Ce résultat est démontré pour les trois modèles de visco-élasticité les plus employés en CND :

Maxwell, Zener et Kelvin-Voigt. La stabilité reste gouvernée uniquement par le domaine de fond, ce qui représente une amélioration significative par rapport au schéma explicite de référence.

Deuxièmement, nous étendons aux modèles visco-élastiques la procédure de stabilisation des schémas explicites, couplée à une méthode de décomposition de domaine. Dans la zone « perturbée », le terme raide est remplacé par un polynôme de Chebyshev de stabilisation dont l'ordre permet d'augmenter le pas de temps admissible de manière proportionnelle. Cette méthode reste explicite en pratique, tout en nécessitant uniquement une résolution locale à l'interface. Les preuves de décroissance d'énergie établissent une condition de stabilité améliorée et contrôlable.

Ces stratégies sont ensuite évaluées en situation réaliste, dans des configurations 2D et 3D représentatives de contrôles ultrasonores dans des matériaux hétérogènes. Les résultats montrent des gains significatifs en pas de temps, en nombre d'itérations et en coût CPU. Dans certains cas, la réduction du temps de calcul atteint un ordre de grandeur tout en reproduisant fidèlement les phénomènes d'atténuation et de réflexion multiple recherchés dans les matériaux micro-structurés.



ALEXANDRE IMPERIALE
Ingénieur-chercheur



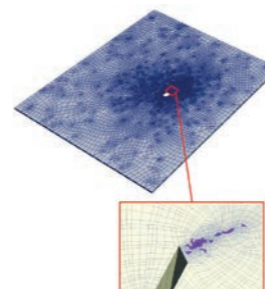
SÉBASTIEN IMPERIALE
Chercheur Inria, école Polytechnique, CNRS

« Nous montrons qu'il est possible de contourner les limitations de certains algorithmes numériques utilisés dans la plateforme CIVA en présence de zones géométriques et/ou de matériaux complexes. »

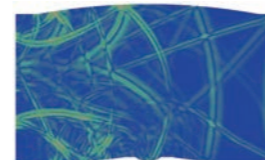
••➤ ALEXANDRE IMPERIALE

« Les méthodes déployées représentent une extension importante des techniques de discrétisation en temps évoluées pour les modèles de propagation d'ondes visco-élastiques. »

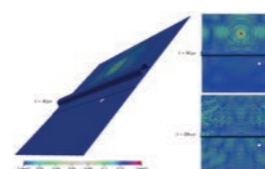
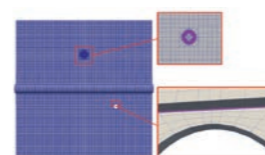
••➤ SÉBASTIEN IMPERIALE



Maillage de plaque perturbée par une fissure



Propagation d'une onde dans un chanfrein de soudure



2D
3D

Temps de calcul un ordre de grandeur inférieur aux approches standards sur la simulation de certaines configurations 2D/3D de contrôle ultrasonore.

••➤ **Cas d'usage, application, transfert**
Modélisation visco-élastique avancée pour l'aide au dimensionnement d'essais ultrasonores haute fréquence sur matériaux hétérogènes et pièces complexes (aéronautique, énergie, pétro-chimie...).

••➤ **Projet et/ou partenariat majeur**
Projet CARNOT CIVAMONT finançant le contrat post-doctoral de Vinuja VASANTHAN sur les méthodes numériques pour le CND ultrasonore. Collaboration avec Sébastien IMPERIALE, chercheur Inria.

••➤ **Publication majeure**
« **Locally implicit and stabilized explicit time schemes for transient visco-elastic wave propagation problems** », V. Vasanthan, A. Imperiale, S. Imperiale, **à paraître dans Journal of Numerical Mathematics**, <https://inria.hal.science/hal-05003766/document>
L'article propose une extension des méthodes localement implicite et explicite stabilisée aux modèles de propagation d'ondes visco-élastiques (Maxwell, Zener, Kelvin-Voigt). Des cas d'application 2D/3D en CND ultrasonore confirment une amélioration significative (en termes de coût de calcul) par rapport aux approches standards utilisées dans la plateforme CIVA.



Jumeau numérique et ingénierie système

Méthodologies

Les jumeaux numériques, un impact tout sauf virtuel

L'utilisation des Jumeaux Numériques (JNs) a été considérablement développée et repose sur des technologies complexes et variées. Cette étude propose une méthodologie qualitative permettant d'évaluer les coûts économiques et environnementaux des JNs. Elle offre aux décideurs des critères actionnables pour les éco-concevoir, sans recourir à des méthodes plus quantitatives.

Les jumeaux numériques sont devenus ces dernières années une technologie clé, reposant sur des innovations variées (IoT, IA, réalité virtuelle, etc.). Pourtant, leur déploiement soulève une question cruciale : **ces outils sont-ils à la fois économiquement viables et écologiquement durables ?**

Leur adoption a débuté en maintenance industrielle, secteur où leur maturité technologique est aujourd'hui la plus avancée. La littérature scientifique montre que la majorité des applications concernent la maintenance prédictive, reposant sur l'IA et le traitement de données volumineuses.

Plus récemment, les technologies de réalité virtuelle et augmentée ont émergé, entraînant un surcoût significatif en termes de ressources matérielles et logicielles.

Cette étude propose une méthodologie qualitative pour classer les JNs de maintenance industrielle, évaluant leurs coûts économiques et environnementaux. Elle aide les décideurs à poser les bonnes questions dès la conception, étape où les données quantitatives manquent souvent. Cette approche est plus accessible que les méthodes très répandues comme l'analyse en cycle de vie (ACV) et offre un cadre simplifié pour l'éco-conception.

Cette méthodologie qualitative s'appuie sur 3 critères clés, classés en 5 niveaux, qui influencent les coûts économiques et environnementaux des JNs (cf. figure 1) :

- **La maturité** : on qualifie de JN mature un JN qui offre des fonctionnalités plus avancées comme la prédiction et la prise de décision améliorée, nécessitant ainsi plus de données,
- **L'acquisition d'équipements** : les coûts environnementaux des JNs dépendent de leur infrastructure matérielle. Un JN utilisant des équipements existants et des capteurs préinstallés a une empreinte plus faible qu'un système nécessitant du matériel neuf,
- **L'expérience utilisateur** : un JN proposant des visualisations basiques, comme des tableaux de bord, aura des impacts économiques et environnementaux plus faibles que celui intégrant des technologies de réalité virtuelle ou augmentée.

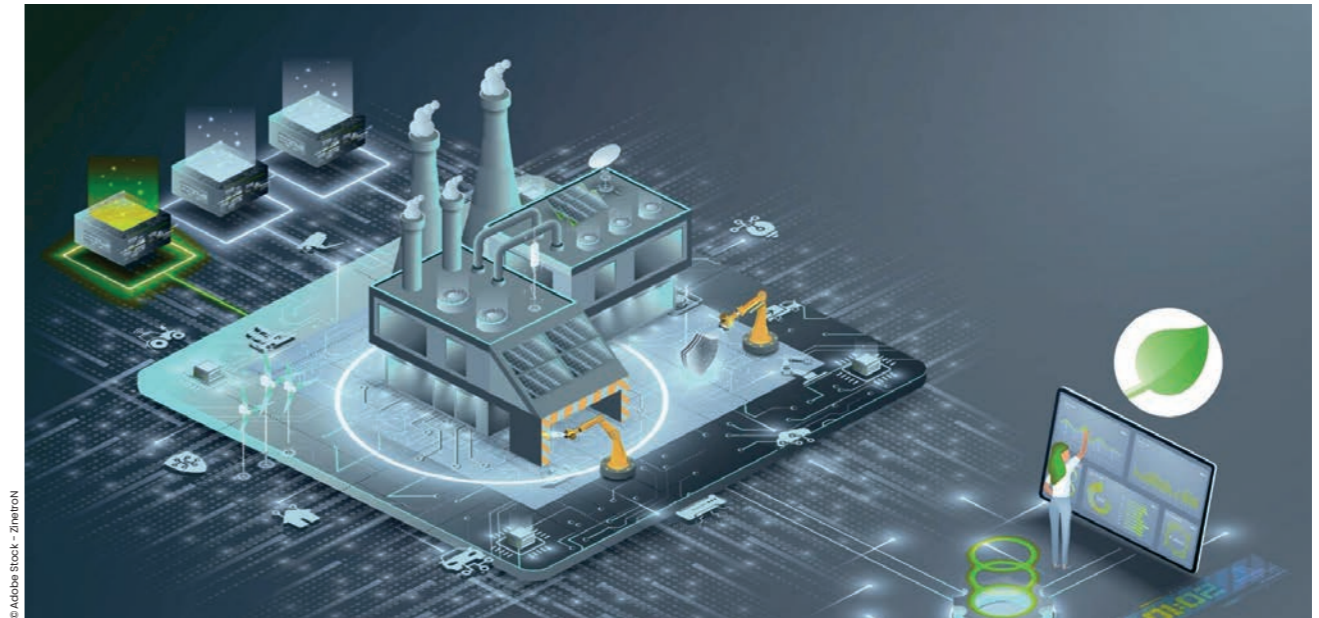
Notre méthode attribue un score global aux JNs en additionnant les niveaux des trois critères décrits précédemment. Un score élevé indique un impact environnemental et économique plus significatif. Nous avons appliqué notre méthode sur trois JNs issus de la littérature : un JN de maintenance prédictive (Cognitive DT), un JN collaboratif pour la maintenance à distance (Collaborative DT) et un JN de prédiction de durée de vie utile restante (DRDT). La figure 2 représente les résultats de l'évaluation de ces trois JNs.



KENZA ELBAROUDI
Apprentie
ingénieure-chercheuse



JEAN-FRANÇOIS BERRÉE
Ingénieur-chercheur



Cette approche, applicable à divers types de JNs, a des limites : elle ne couvre pas la gestion des données, ni le cycle de vie complet et ne distingue pas les coûts économiques et environnementaux, qui n'évoluent pas toujours dans le même sens.

En conclusion, notre méthodologie fournit une compréhension initiale des impacts environnementaux et économiques des JNs et elle nécessite un développement et un raffinement supplémentaires pour devenir plus précise.

	Maturity	Equipment	UX
L1	Mirror	No new equipment	Basic visualizations
L2	Shadow	IoT	Advanced visualizations
L3	Control	One simple user terminal	3D modélisation
L4	Cognitive	One augmented user terminal	Extended Reality (XR)
L5	Collaborative	Multiple user terminals	Gamification

Figure 1 - Grille utilisée pour évaluer un JN.

« Cette approche aide les décideurs à se poser les bonnes questions et leur fournit un point de départ plus accessible et plus rapide pour l'écoconception. »

••• JEAN-FRANÇOIS BERRÉE

80%

Près de 80 % de l'impact environnemental des produits et technologies numériques provient d'équipements physiques tels que les appareils IoT et les écrans d'ordinateur.

Publication majeure



« Towards A Methodological Framework For Early Qualitative Assessment Of The Ecological And Economic Costs Of Digital Twins In Industrial Maintenance »
<https://doi.org/10.3850/978-981-94-3281-3-ESREL-SRA-E2025-P5071-cd>
 K. Elbaroudi, N. Julien, J-F. Berrée, R. Levailant, A. Piel. Cette étude a été présentée aux conférences jointes European Safety and Reliability (ESREL) and Society for Risk Analysis Europe (SRA-E) en Norvège en juin 2025.

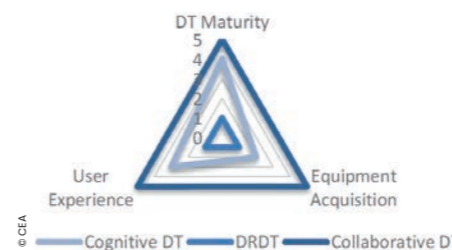


Figure 2 - Représentation de l'évaluation des 3 exemples de JN.

Jumeau numérique et ingénierie système

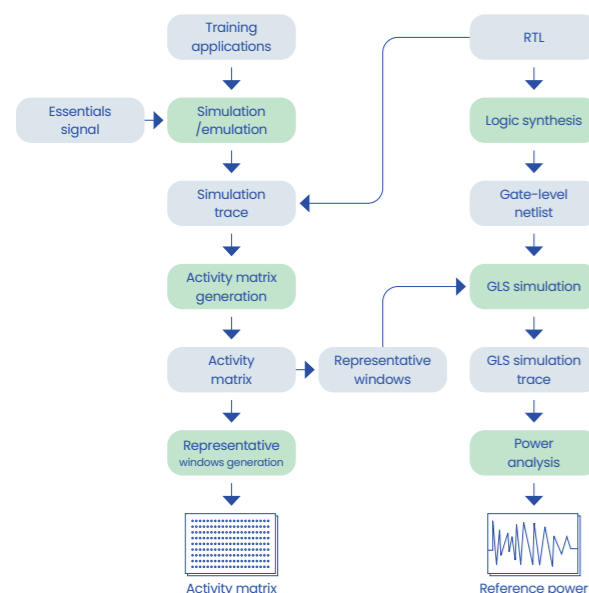
Méthodologies

Une méthodologie basée sur l'apprentissage automatique pour une modélisation rapide et efficace de la consommation de puissance des architectures numériques

Cette étude introduit une méthodologie permettant d'accélérer de manière significative la modélisation de consommation de puissance assistée par l'IA. En réduisant automatiquement les traces RTL (Register-transfer-level) grâce à la sélection de fenêtres représentatives via des techniques de clustering, elle permet d'obtenir des modèles précis tout en diminuant fortement le volume de données et le temps de génération nécessaires.



Figure 1 : Flot de génération du jeu de données d'entraînement réduit.



La méthodologie proposée (Figure 1) vise à accélérer la modélisation de la consommation de puissance pour les architectures numériques en combinant apprentissage automatique et réduction drastique du volume de données d'entraînement. La création de tels modèles repose sur des jeux de données associant des traces au niveau de transfert de registres décrivant l'activité interne du circuit à des valeurs de consommation de référence obtenues via des outils commerciaux d'analyse de puissance au niveau des portes logiques.

Cependant, ce processus demeure limité par deux goulots d'étranglement majeurs : la lourdeur des simulations au niveau des portes logiques et le coût élevé de la génération des profils de consommation correspondants.

Pour dépasser ces limitations, la méthodologie introduit un mécanisme automatisé de sélection de fenêtres représentatives, fondé sur un clustering « K-means ». Les traces au niveau du transfert de registres sont d'abord segmentées en fenêtres de taille fixe, puis

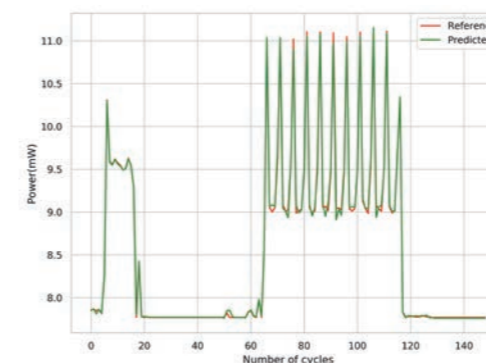


Figure 2 : Exemple de comparaison entre la puissance de référence et la puissance prédite pour un AES masqué.

agrégées afin de permettre un clustering robuste sur des séries temporelles complexes. Les fenêtres ainsi identifiées permettent de restreindre l'analyse à un sous-ensemble réduit mais informatif de segments d'exécution, sur lesquels les simulations au niveau des portes logiques et les calculs de puissance sont concentrés. Cela permet de réduire fortement le volume de données à traiter sans compromettre la diversité comportementale nécessaire à l'apprentissage.

Les résultats expérimentaux, obtenus sur un cœur RISC-V Rocket et un AES masqué, mettent en évidence des gains significatifs : jusqu'à **28x** de réduction du temps d'analyse de puissance et jusqu'à **49x** de réduction du temps d'entraînement des modèles. Malgré cette compression des données, l'erreur moyenne de prédiction reste autour de **5%**. La Figure 2 illustre, à titre d'exemple, une comparaison cycle par cycle d'un segment de la puissance prédite (en vert) et de la puissance de référence (en rouge) pour un AES masqué.

Jusqu'à 28x d'accélération de la phase de génération de données et 49x pour l'entraînement des modèles de puissance, tout en conservant une erreur de prédiction autour de 5%.

28x
49x

« En optimisant la sélection des données d'apprentissage, cette approche assistée par l'intelligence artificielle réduit considérablement les besoins en simulation tout en préservant la précision des modèles de consommation de puissance, ouvrant la voie à une modélisation plus rapide et plus scalable »

• CAALIPH ANDRIAMISAINA

• **Cas d'usage, application, transfert**
Génération accélérée de modèles de puissance pour le cœur RISC-V Rocket grâce à la réduction automatique des fenêtres représentatives. Application aux systèmes cryptographiques tels que l'AES masqué, permettant une analyse énergétique rapide dans des environnements de sécurité.

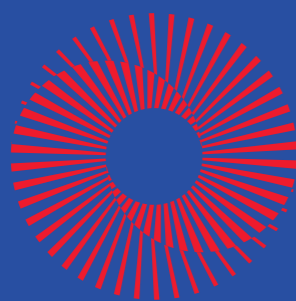
• **Brevet**
« Procédé de construction et d'entraînement d'un détecteur de la présence d'anomalies dans un signal temporel, dispositifs et procédé associés ». Dépôt France : FR2014188 (réf CEA : DD20704)

• **Projet et/ou partenariat majeur**
Ces travaux sont alignés avec les activités d'IA pour l'EDA du CEA-List, visant à intégrer l'IA pour accélérer et optimiser la conception d'architectures matérielles. Cette approche est actuellement mise en œuvre dans le projet européen TRISTAN, et sera également exploitée dans le projet européen RIGOLETTO et dans le projet national CADabriA du programme PHOENIX. Elle est par ailleurs utilisée dans le projet industriel avec SILVACO.

• **Publication majeure**
« A Methodology for Fast and Efficient ML-based Power Modeling » C. Andriamisaina, K. Trabelsi, P-G. Le Guay, ICCD 2024. <https://doi.org/10.1109/ICCD63220.2024.00109>



02



Les technologies de l'intelligence artificielle

Les travaux menés au CEA-List portent sur le développement de méthodes avancées d'intelligence artificielle couvrant l'IA générative, le traitement automatique des langues et la vision par ordinateur, avec un objectif commun de robustesse, d'efficacité et de confiance en conditions réelles.

En IA générative, le CEA-List a développé un modèle fondé sur la génération augmentée de récupération pour la réponse à des questions visuelles portant sur des entités nommées, surpassant l'état de l'art. En vision, la découverte et la segmentation d'objets sans annotations humaines sont abordées par le développement d'une méthode de distillation croisée exploitant la complémentarité des données 2D et 3D, ainsi que par l'analyse sémantique de scènes 3D à partir de requêtes en langage naturel pour la robotique et la réalité augmentée.

En IA embarquée et distribuée, des approches basées sur des graphes d'événements asynchrones exploitent les caméras événementielles pour la détection de mouvements à très faible latence, tandis que l'intégration de l'analyse au plus près des capteurs ouvre la voie à des systèmes d'acquisition intelligents pour l'inspection des structures. Les enjeux d'apprentissage collaboratif et adaptatif sont traités par des méthodes fédérées décentralisées pour l'adaptation de domaine multi-source et par des algorithmes limitant l'oubli catastrophique et la dérive de concept, appliqués notamment à la gestion prédictive et confidentielle des bornes de chargement électriques.

Enfin, la confiance et la sûreté des IA sont adressées par l'analyse des biais des grands modèles de langue, la vérification formelle des réseaux de neurones et l'estimation de l'incertitude dans les simulations Monte Carlo assistées par l'IA.

30 IA générative, NLP, vision

- • 30 MiRAG : génération augmentée de récupération multi-niveau pour répondre à des questions visuelles
- • 32 Découvrir le monde sans annotations avec xMOD : collaboration entre vision 2D et 3D
- • 34 Analyse de scène 3D à partir de requêtes en langage naturel

36 IA embarquée et distribuée

- • 36 Détection de mouvement à faible latence avec des graphes d'événements
- • 38 Vers des systèmes d'acquisition intelligents
- • 40 L'adaptation décentralisée de domaine multi-source dans le cadre d'une approche fédérée : pour un apprentissage collaboratif sans serveur
- • 42 Apprentissage adaptatif pour contrer l'oubli catastrophique et la dérive de concept dans un contexte fédéré : application à une gestion optimisée des bornes de chargement électriques alliant prédiction et confidentialité

44 Confiance et sûreté des IA

- • 44 Analyser et réduire les biais politiques dans les grands modèles de langue
- • 46 La vérification des réseaux de neurones : un défi à relever
- • 48 Incertitude dans les simulations Monte-Carlo aidées par l'IA

Les technologies de l'intelligence artificielle

IA générative, NLP, vision

MiRAG : génération augmentée de récupération multi-niveau pour répondre à des questions visuelles

Le CEA-List a développé le modèle MiRAG pour la tâche de réponse à des questions visuelles à propos d'entités nommées, en y appliquant pour la première fois une approche d'intelligence artificielle générative fondée sur la génération augmentée de récupération (RAG). Les évaluations menées sur le benchmark de référence ViQuAE ont montré la supériorité de MiRAG par rapport aux approches existantes dans le domaine.



OMAR ADJALI
Post-doctorant



OLIVIER FERRET
Directeur de recherche et expert fellow



HERVÉ LE BORGNE
Directeur de recherche et expert senior

La réponse à des questions visuelles à propos d'entités nommées (KVQAE) a récemment fait l'objet d'une attention particulière en tant que tâche de référence pour évaluer les capacités des systèmes de recherche d'information à comprendre en profondeur les informations visuelles et textuelles. Alors que dans la tâche standard de question-réponse visuelle (VQA), la réponse aux questions peut se faire uniquement à partir des images, la KVQAE impose la recherche des réponses dans une base de connaissances externes constituée à la fois de textes et d'images. Comme l'illustre la figure en entête, un système de KVQAE prend en entrée une question textuelle (combien de symphonies a-t-il composé ?) et une image (une photo de Johannes Brahms) et doit chercher la réponse à la question posée dans un ensemble de documents mêlant texte et images (des pages Wikipédia).

Dans ce contexte, le CEA-List a proposé le modèle MiRAG, s'appuyant sur le paradigme du RAG pour résoudre la tâche KVQAE, en associant un modèle de recherche d'information et un modèle de génération des réponses. La figure principale illustre les principales étapes de cette approche. L'idée de base est d'effectuer la recherche à différents niveaux de granularité (entité et passage) à partir d'une requête multimodale avec des entrées textuelles et visuelles. Cette stratégie

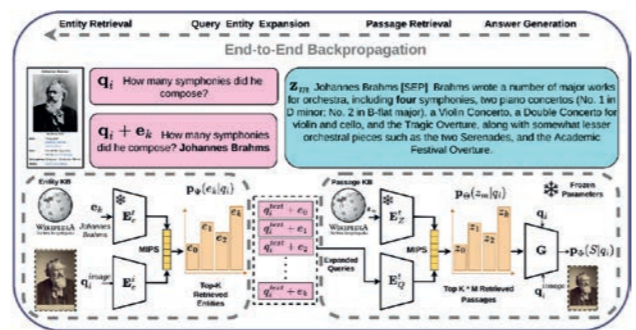


Figure principale : Vue d'ensemble du modèle MiRAG

d'extraction à plusieurs niveaux affine progressivement la requête avant de générer des réponses à l'aide d'un modèle de génération.

La recherche est d'abord effectuée à un premier niveau, celui des entités, assimilable au niveau des documents, afin d'identifier un ensemble d'entités candidates pertinentes pour la requête. Les entités ainsi récupérées sont ajoutées à la requête avant d'effectuer une nouvelle recherche à un niveau plus fin, celui du passage. L'ajout d'entités aux requêtes permet de spécifier de façon plus précise ces dernières, ce qui

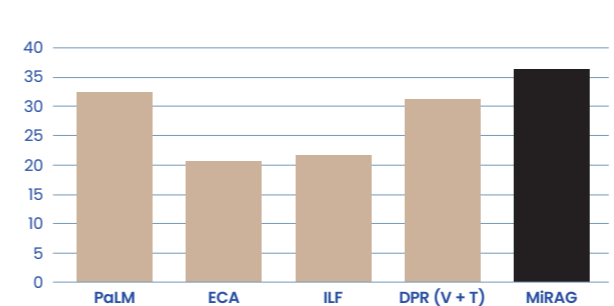


Figure 1 : Résultats de MiRAG comparés à des méthodes de référence sur le benchmark ViQuAE (mesure : exact match).

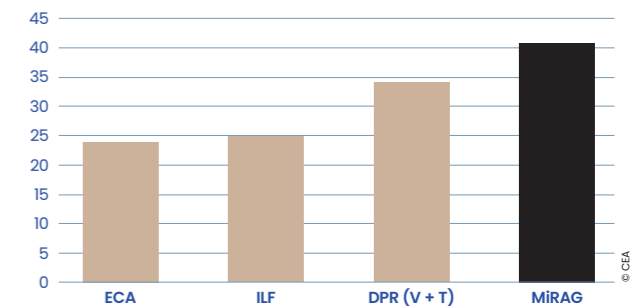


Figure 2 : Résultats de MiRAG comparés à des méthodes de référence sur le benchmark ViQuAE (mesure : F1).

aboutit à la sélection de passages plus pertinents et fournit un contexte supplémentaire au générateur de réponses. Formellement, l'objectif est d'apprendre la probabilité $p(a|q,z,e)$ de générer une réponse a conditionnée par la requête q , un passage extrait z et l'entité extraite correspondante e .

Cet apprentissage est réalisé de bout-en-bout en associant à la fois le modèle de recherche d'information et le modèle de génération des réponses.

La comparaison de MiRAG à des modèles de référence associant texte et image pour la tâche de KVQAE (cf. Figures 1 et 2), en l'occurrence ECA, ILF et DPR[V+T], montre que conjuguer une approche RAG et la stratégie de recherche à plusieurs niveaux implémentée par MiRAG permet une meilleure exploitation de la synergie entre les deux modalités.

Par ailleurs, alors qu'un très gros modèle génératif textuel tel que PALM, issu de Google, peut surclasser les modèles multimodaux considérés, l'exploitation conjointe du texte et de l'image par MiRAG permet de dépasser nettement les performances de PALM avec 138 fois moins de paramètres.

Projet et/ou partenariat majeur

Le CEA-List a développé et appliqué ces travaux dans le cadre du projet ANR MEERQAT (2020-2024), réunissant le CEA, le LISN, l'IRT ainsi que l'IRISA. Le projet se focalise sur la représentation multimodale (texte et image) des entités et l'exploitation de ces représentations pour des tâches de question-réponse.

Publication majeure



« Multi-Level Information Retrieval Augmented Generation for Knowledge-based Visual Question Answering ». <https://aclanthology.org/2024.emnlp-main.922/> Omar Adjali, Olivier Ferret, Sahar Ghannay and Hervé Le Borgne (2024). 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), p. 16499-16513, Miami, Florida, USA.

Les technologies de l'intelligence artificielle

IA générative, NLP, vision

Découvrir le monde sans annotations avec xMOD : collaboration entre vision 2D et 3D



Découvrir et localiser des objets en 3D sans aucune annotation humaine est un défi majeur en vision par ordinateur. xMOD est une méthode innovante de distillation croisée qui fait collaborer caméras (2D) et capteurs LiDAR (3D). En exploitant les indices de mouvement présents dans l'image, l'IA apprend seule à segmenter son environnement, offrant des performances au-delà de l'état de l'art.

La détection d'objets classique repose sur des données massivement annotées manuellement, un processus coûteux et peu évolutif. L'approche par la « découverte d'objets » vise à s'affranchir de cette supervision humaine. Appliqué à l'analyse d'images 2D, le principe est de considérer un groupe de pixels se déplaçant avec une trajectoire cohérente comme un seul et même objet. Cette approche restait jusqu'ici peu explorée dans le domaine de la 3D, en raison de la nature éparses et incomplètes des nuages de points LiDAR.

Pour répondre à ce défi, le CEA-List a proposé une première base de référence pour la découverte d'objets 3D utilisant le mouvement 2D (DIOD-3D), et conçu xMOD, un cadre d'entraînement par distillation croisée. xMOD repose sur une architecture « enseignant-élève » symétrique : un modèle expert 2D guide l'apprentissage du modèle 3D, et réciproquement. Cette double interaction exploite la complémentarité des modalités de données et permet de pallier les faiblesses de chacune d'elles : la caméra aide le LiDAR lorsque les points sont trop rares, et le LiDAR aide la caméra dans des conditions visuelles difficiles (nuit, manque de texture...).

Une autre innovation réside dans la **complétion de scène** comme tâche prétexte pour la 3D. Contrairement à la reconstruction simple, le modèle est entraîné à prédire des nuages de points plus denses, ce qui l'aide à appréhender la géométrie des objets, malgré la pauvreté des données d'entrée.

Avec xMOD, l'étiquetage de millions de scènes peut être réalisé sans aucune intervention humaine, avec une précision augmentée par l'utilisation conjointe des modalités caméra et LiDAR. L'approche a été évaluée sur trois jeux de données de référence : synthétiques (TRIP-PD) et réelles (KITTI, Paymo). Le modèle surpasse l'état de l'art de la découverte d'objets 2D avec des gains allant de **+8,7** (KITTI) à **+15** points (Waymo) de score F1.

La stratégie de fusion tardive (combinant les prédictions 2D et 3D lors de l'inférence) améliore encore la fiabilité, notamment pour les objets situés à moyenne distance (10-30 mètres).

Ces avancées en découverte d'objets ouvrent la voie à une conception plus rapide et plus fiable de modèles de perception 3D indispensables aux systèmes de conduite automatisée.

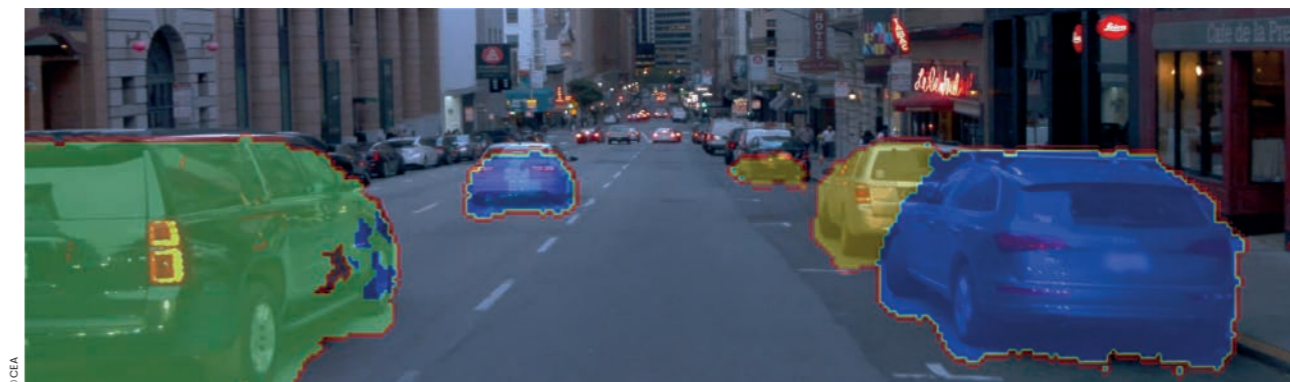


Figure 1 : Masques d'instances prédits par xMOD (2D+3D).

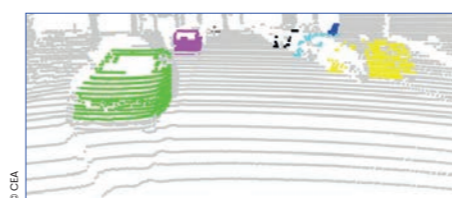


Figure 2 : Visualisation 3D des prédictions produites par xMOD (2D+3D).

« Embarquer les traitements IA dans les systèmes et appareils intelligents »

•• SAAD LAHLALI

+15,1

gain maximal de score F1@50 obtenu par l'approche xMOD par rapport à l'état de l'art de la découverte d'objets 2D.

•• Applications
L'algorithme, principalement destiné ici aux **véhicules autonomes** et à la **robotique mobile**, peut être généralisé aux systèmes nécessitant une perception robuste et évolutive sans étiquetage manuel massif. Il est particulièrement pertinent pour améliorer la robustesse dans des conditions dégradées (nuit pour la caméra, objets peu réfléchissants pour le LiDAR) grâce à la redondance multimodale.

•• Projet
DeepGreen ANR-23-DEGR-0001. <https://deepgreen.ai/>

•• Publication majeure
« Cross-Modal Distillation for 2D/3D Multi-Object Discovery from 2D motion » <https://arxiv.org/abs/2503.15022> Lahlali et al., CVPR 2025.

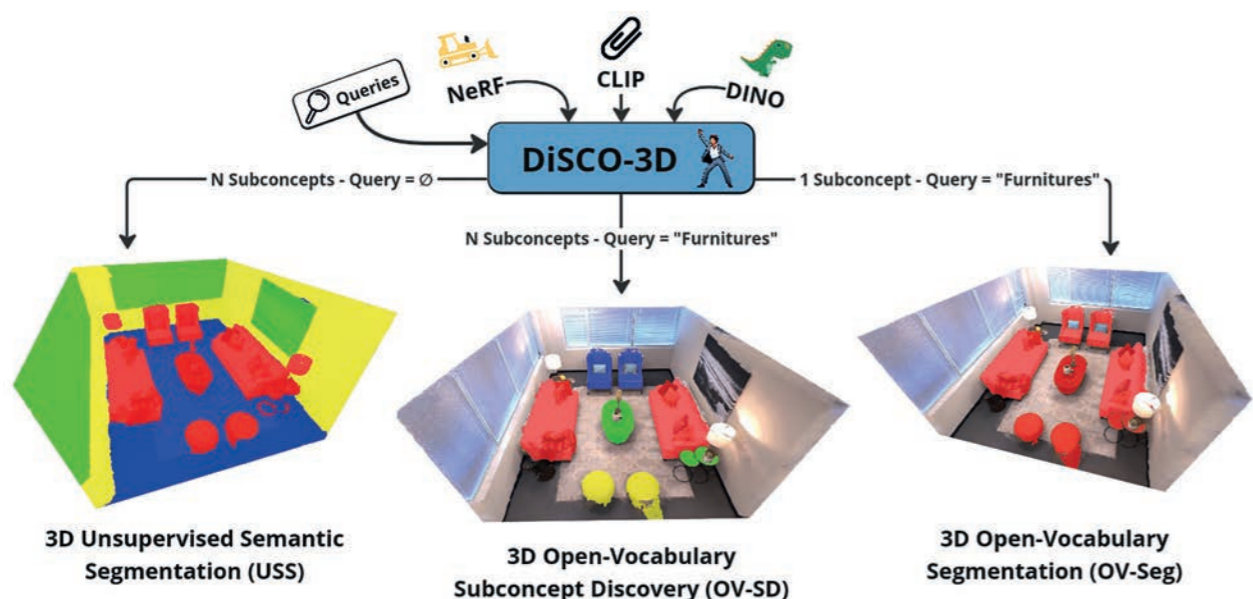


Les technologies de l'intelligence artificielle

IA générative, NLP, vision

Analyse de scène 3D à partir de requêtes en langage naturel

La méthode de segmentation sémantique DISCO-3D permet de découvrir, dans une scène 3D, les éléments correspondant à des sous-concepts sémantiques d'une requête utilisateur exprimée en langage naturel. Cette méthode offre, grâce à une **approche basée NeRF***, une compréhension de haut niveau des scènes pour des applications de robotique ou de réalité augmentée par exemple.



DISCO-3D unifie la segmentation sémantique 3D non supervisée et en vocabulaire ouvert afin de découvrir des sous-concepts sémantiques adaptés au contenu de la scène.

* NeRF : Les Neural Radiance Fields sont une technologie à l'état de l'art qui permet de reconstruire avec un réseau de neurones des scènes 3D à partir d'images 2D.
** LeRF : Les Language Embedded Radiance Fields étendent les NeRF en associant une information sémantique à chaque point de l'espace.

Les méthodes de segmentation sémantique 3D actuelles proposent soit d'identifier les objets correspondant à un unique concept sémantique recherché par l'utilisateur (segmentation en vocabulaire ouvert ou OV-Seg), soit de s'adapter au contenu de la scène en découvrant plusieurs concepts sémantiques (segmentation sémantique non supervisée ou USS).

DISCO-3D est la première méthode capable d'unifier les deux paradigmes, en abordant le problème plus large de la découverte de sous-concepts sémantiques en vocabulaire ouvert (OV-SD). L'approche proposée vise à découvrir les différents sous-concepts sémantiques de la scène 3D pertinents vis-à-vis d'une requête en langage naturel (figure 1).

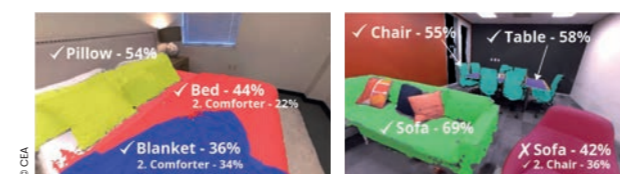


Figure 1 : segmentation des sous-concepts découverts pour la requête « sleep » (image de gauche) et « furniture » (image de droite).

L'architecture de DISCO-3D (figure 2) comporte deux modules. Le premier réalise la tâche d'OV-Seg pour identifier les zones de la scène ne correspondant pas à la requête de l'utilisateur (fond). Le second module réalise l'USS en forçant l'un des segments à se superposer au fond trouvé. Cette supervision assure ainsi que les autres segments découverts par l'USS correspondent à des sous-concepts sémantiques pertinents pour la requête.

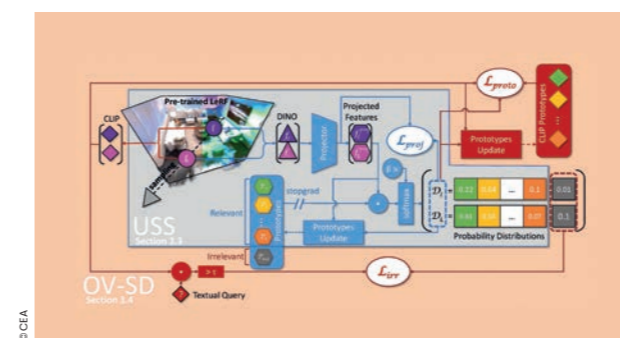


Figure 2 : Présentation de DISCO-3D pour un champ de caractéristiques (features) LeRF**.

L'efficacité de la méthode est démontrée sur des scènes variées avec différentes requêtes utilisateur (figure 3).



Figure 3 : Évaluation qualitative de DISCO-3D pour différentes requêtes.

Enfin, les requêtes étant définies en langage naturel, DISCO-3D peut facilement être intégré comme outil d'une IA agentique, ouvrant les portes de l'analyse de scène 3D à l'aide d'un grand modèle de langage (LLM).

- • • **Domaines d'application**
Reality Capture et robotique autonome

- • • **Brevet**
« Méthode de découverte automatique de sous-concepts sémantiques dans une scène », D. Petit, S. Bourgeois, V. Gay-Bellile, F. Chabot, numéro de dépôt FR2501675

- • • **Publication majeure**
« DISCO-3D : Discovering and segmenting Sub-Concepts from Open-vocabulary queries in NeRF », Dorian Petit, Steve Bourgeois, Vincent Gay-Bellile, Florian Chabot, Loïc Barthe, The IEEE/CVF International Conference on Computer Vision (ICCV) 2025, Honolulu, Hawaii, USA, pp. 20043-20052 https://openaccess.thecvf.com/content/ICCV2025/html/Petit_DISCO-3D_Discovering_and_Segmenting_Sub-Concepts_from_Open-vocabulary_Queries_in_ICCV_2025_paper.html

- • • Ce travail a été rendu possible grâce à l'utilisation du supercalculateur FactoryIA du CEA-List, soutenu par le Conseil régional d'Île-de-France.

Les technologies de l'intelligence artificielle

IA embarquée et distribuée

Détection de mouvement à faible latence avec des graphes d'événements

Les caméras événementielles permettent une détection de mouvement à très faible latence. La méthode proposée, à base de graphe d'événements asynchrone, tire parti de la haute résolution temporelle de ces caméras afin de détecter des mouvements avec une **latence de seulement 50 µs** tout en **réduisant le nombre d'opérations par 48** par rapport à l'état de l'art.

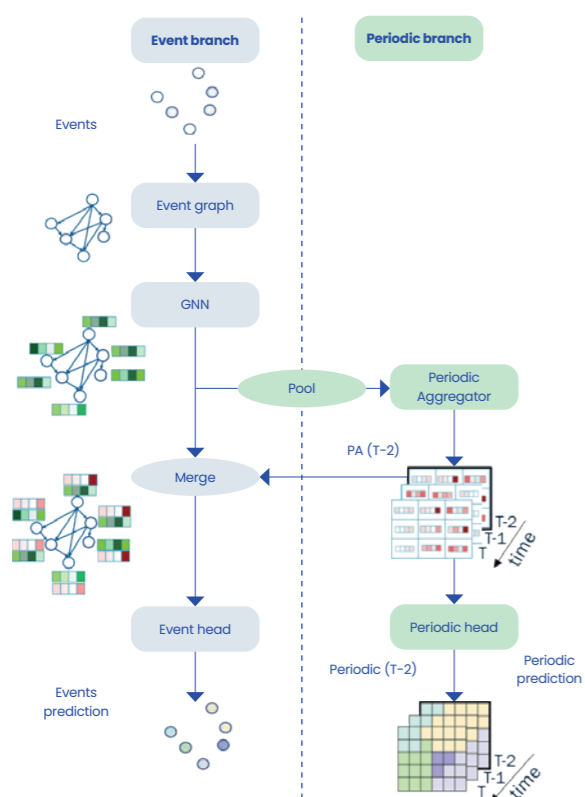


MANON DAMPFHOFFER
Ingénieure-chercheuse



CHRISTOPH POSCH
CTO PROPHESEE

Figure 1: Architecture proposée (HUGNet et Agrégateur périodique).

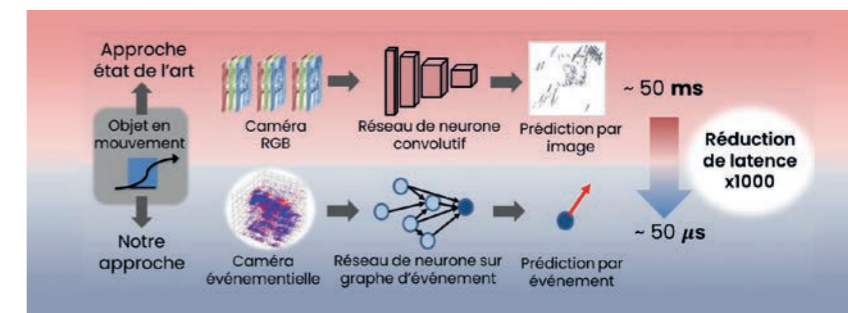


Les systèmes de vision par ordinateur nécessitent une haute précision temporelle, notamment pour la navigation autonome ou la stabilisation vidéo. Les méthodes traditionnelles, basées sur des images RGB, sont limitées par leur coût calculatoire élevé et la fréquence d'acquisition des images. Les caméras événementielles offrent une alternative prometteuse grâce à leur fonctionnement asynchrone et leur haute résolution temporelle. Cependant, les méthodes actuelles reposant sur des réseaux convolutifs recréent des images à partir des événements, perdant ainsi leur avantage en termes de latence et restant trop coûteuses pour des systèmes embarqués.

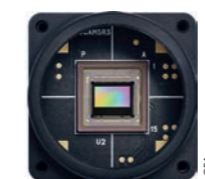
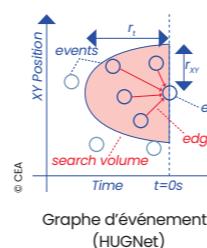
Les graphes d'événements, interprétés par des réseaux de neurones graphes, permettent des prédictions par événement avec une latence minimale. La méthode précédemment développée au laboratoire (HUGNet) a montré que des graphes dirigés mis à jour à-la-volée pouvaient réduire encore cette latence, mais au prix d'une perte de précision, faute d'agrégation de contexte global. Pour dépasser cette limitation, nous proposons une nouvelle architecture de réseaux de neurone, combinant la branche asynchrone avec une branche périodique d'agrégation. Cette branche périodique utilise un réseau convolutif récurrent sur des images d'événements passés, permettant d'extraire un contexte global de la scène, améliorant la qualité de la prédiction sans en augmenter la latence.

« Notre approche ouvre la voie à des capteurs intelligents très faible latence et économes en énergie pour la vision embarquée. »

MANON DAMPFHOFFER

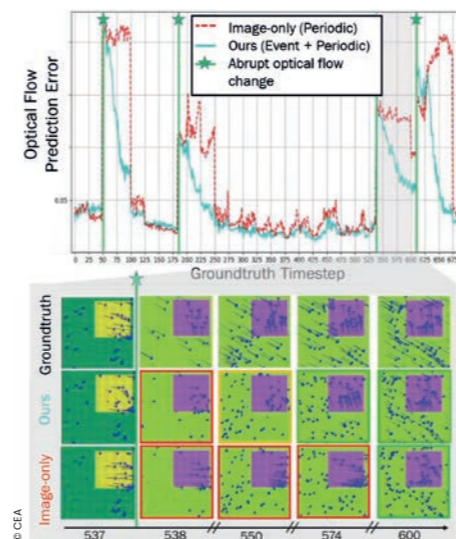


Le traitement par graphe d'événements permet une réduction de la latence de prédiction par 1000 par rapport à un traitement par image.



Capteur événementiel EVK4 HD produit par PROPHESEE

Notre approche réduit la latence de détection des changements de vitesse, avec une prédiction par événement en **50 µs** sur matériel embarqué, soit **1000 fois plus rapide** que les solutions classiques. De plus, elle nécessite **17 fois moins de paramètres** et **48 fois moins d'opérations par seconde**, tout en maintenant une précision compétitive.



Diminution de la latence de détection d'un changement de vitesse.

Cas d'usage, application, transfert
L'algorithme sera intégré dans un imageur événementiel intelligent 3 couches très faible consommation (<10 mW) fabriqué en collaboration avec PROPHESEE et STMicroelectronics. En plus de la prédiction de flux optique, l'algorithme peut permettre la reconnaissance automatique de motifs, tels que des gestes ou la lecture labiale.

Brevet
DD24590ST - En cours de dépôt.
DD22912ST - Déposé.

Projet et/ou partenariat majeur
ANR IRT Nanoélec (programme Smart Imager), en partenariat avec PROPHESEE.

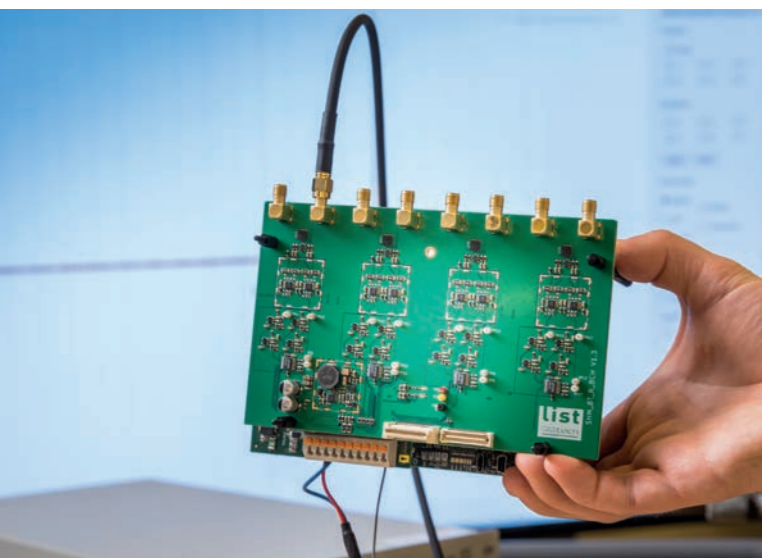
Publication majeure
« Graph Neural Network Combining Event Stream and Periodic Aggregation for Low-Latency Event-Based Vision ».
M. Dampfhofer, T. Mesquida, D. Joubert, T. Dalgaty, P. Vivet and C. Posch (2025).
In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
<https://doi.org/10.1109/CVPR52734.2025.00648>
Sélectionné "Highlights" de la conférence (rang A*).

Les technologies de l'intelligence artificielle

IA embarquée et distribuée

Vers des systèmes d'acquisition intelligents

L'essor du calcul embarqué et de l'intelligence artificielle permet désormais d'intégrer l'analyse directement au sein des systèmes d'acquisition. Ces technologies ouvrent de nouvelles perspectives pour l'assistance à l'inspection en contrôle non destructif et pour l'automatisation du contrôle santé des structures (SHM) sur structures instrumentées.



Carte d'acquisition analogique intégrée au système GERONIMO.

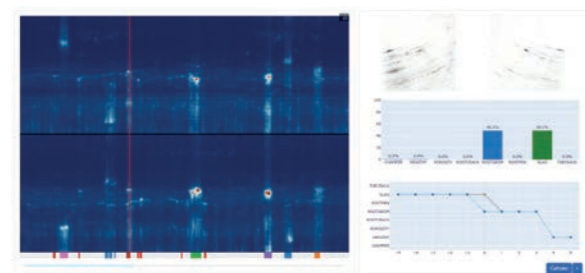


Figure 1 : À gauche, inspection complète d'une soudure par les deux sondes (vue du dessus) ; à droite, imagerie TFM par les deux sondes, résultat de la prédiction courante, comparaison des prédictions avec la vérité expert.



CLÉMENT FISHER
Ingénieur-chercheur



ROBIN GUYON
Ingénieur-chercheur

Cette évolution vers des systèmes d'acquisition intelligents se manifeste différemment selon les types d'inspection. En contrôle non destructif (CND), l'objectif est d'assister l'opérateur en fournissant une analyse fiable pour guider l'inspection et la prise de décision. En SHM, les capteurs génèrent en continu des volumes importants de données. L'enjeu est d'automatiser l'analyse de ces données au niveau des capteurs afin de limiter la transmission d'information. Grâce à notre maîtrise complète du hardware, nous avons la capacité d'intégrer les traitements et les modèles d'IA dans les systèmes d'acquisition. Des outils comme le *framework* AIDGE, dédié à la manipulation, la conversion et l'optimisation des réseaux de neurones pour cibles embarquées, ou encore le système d'acquisition GERONIMO pour la mesure d'ondes guidées, permettent de déployer des solutions complètes, du capteur jusqu'au diagnostic. Deux preuves de concept issues des travaux du CEA-List illustrent ces possibilités.

Pour l'inspection des soudures par ultrasons^[1], un modèle de type ConvNeXt, allégé pour répondre aux contraintes de l'embarqué, a été implémenté pour l'analyse d'images TFM. Entraîné sur des données expérimentales, il permet une classification sans faux négatif, critère essentiel en contexte industriel, tout en identifiant 87 % des données saines dans les conditions nominales d'inspection. La robustesse a été validée en modifiant les conditions d'inspection (décalage des sondes lors de l'inspection) ainsi que la géométrie de la structure (tuyau, plaque, variations d'épaisseur). Le modèle a été optimisé *via* AIDGE et déployé efficacement sur une architecture NXP i.MX8 (figure 1), base de nos différents systèmes d'acquisition, atteignant en inférence une fréquence d'environ 10 Hz.

« Grâce à l'analyse embarquée des ondes guidées, les données de mesure deviennent des indicateurs clairs de l'état de la structure, qui permettent une surveillance constante. »

••• ROBIN GUYON

Compression embarquée : jusqu'à 10 000× moins de données (3 valeurs flottantes au lieu de plusieurs kB)

÷10000

Dans le cas du SHM par ondes guidées^[2], l'objectif est d'automatiser l'ensemble de la chaîne, de l'acquisition jusqu'à la prise de décision. Ici, nous présentons un CNN dédié au traitement d'images pour réaliser l'analyse des images *delay-and-sum* (DAS, figure 2). Bien que cette méthode d'imagerie qualitative ne soit pas la plus performante en termes de résolution, elle reste largement utilisée dans l'état de l'art et constitue donc un cas d'étude pertinent. Le modèle est entraîné sur des données simulées avec le logiciel CIVA, puis ajusté par transfert *learning* sur quelques mesures expérimentales, permettant ainsi de localiser mais aussi de dimensionner les défauts. Converti *via* AIDGE et embarqué sur le système d'acquisition GERONIMO (figure 3), le modèle réduit le volume des données issues des capteurs de plusieurs kilo-octets à seulement trois valeurs flottantes. Cette compression réduit considérablement les communications et rend possible une surveillance continue, y compris sur des structures fortement instrumentées. Les travaux se poursuivent pour enrichir l'offre de méthodes de traitements, telles que la tomographie par ondes guidées, et d'IA dans le système GERONIMO et sa version OPTOGERO dédiée à la mesure par réseaux de Bragg^[3].

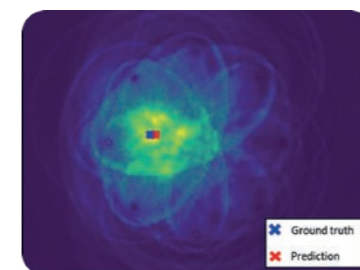


Figure 2 : Imagerie DAS avec prédiction de la position du défaut par le modèle.

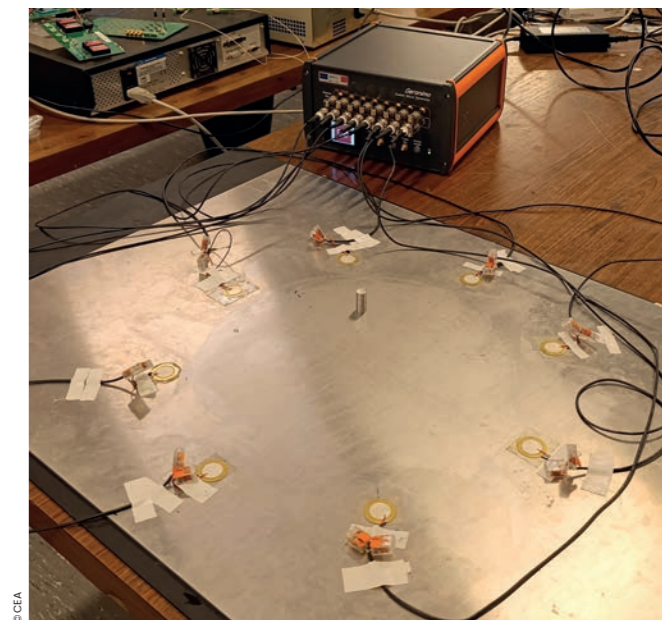


Figure 3 : Plaque instrumentée avec le système GERONIMO.

••• Cas d'usage, application, transfert
GERONIMO : partenariat UGE et licence à Capturia pour la commercialisation

••• Projet et/ou partenariat majeur
Find, Renaissance Fusion, Alstom RSHM, DeepGreen, Multimod'AIR.

••• Publications majeures

[1] « Embedded Artificial Intelligence in Guided Wave SHM system: Signal processing, and data analysis ». C. Fisher, A. Recoquillay, B. Chapuis, and P. Calmon, *e-Journal of Nondestructive Testing*, vol. 30, no. 06, Jun. 2025, <https://doi.org/10.58286/31305>

[2] « Towards embedded AI models for welding defect detection in pipes ». R. Guyon, M. Newson, C. Fisher, R. Miorelli, and D. Roué, *in 2025 IEEE Sensors Applications Symposium (SAS)*, Jul. 2025, pp. 1–6, <https://doi.org/10.1109/SAS65169.2025.11105144>

[3] « Embedded passive guided wave tomography. Application to corrosion monitoring in multilayered pipes ». A. Recoquillay et al., *e-Journal of Nondestructive Testing*, vol. 29, no. 07, Jul. 2024, <https://dx.doi.org/10.58286/29855>

Les technologies de l'intelligence artificielle

IA embarquée et distribuée

L'adaptation décentralisée de domaine multisource dans le cadre d'une approche fédérée : pour un apprentissage collaboratif sans serveur

Le travail présenté introduit De-FedDaDiL, une méthode totalement décentralisée pour l'adaptation de domaine multisource (MSDA). En s'appuyant sur les barycentres de Wasserstein et le *dataset dictionary learning*, elle permet d'adapter des modèles à un domaine cible non étiqueté sans serveur central, tout en améliorant robustesse, scalabilité et confidentialité.



FRED NGOLE MBOULA
Ingénieur-chercheur et expert

Comprendre le défi : s'adapter à des données hétérogènes et dispersées

L'adaptation de domaine multi-source (MSDA) répond à un enjeu central : permettre à un modèle d'exploiter plusieurs sources de données hétérogènes pour généraliser vers un domaine cible non annoté. Dans de nombreuses applications industrielles, les données sont non seulement différentes d'un site à l'autre, mais aussi souvent impossibles à centraliser. Le décalage de distribution entre sources et cible dégrade fortement les performances des modèles classiques, qui supposent une homogénéité rarement observée dans la réalité.

Les limites du fédéré classique : un serveur central qui fragilise tout le système

Les approches fédérées récentes ont tenté d'adapter la MSDA au contexte distribué en préservant la confidentialité des données. Toutefois, elles reposent presque toutes sur un serveur central chargé d'agrèger et de synchroniser les modèles locaux. Ce fonctionnement introduit plusieurs faiblesses : dépendance à un point unique, risques accrus en cybersécurité et coût de communication élevé.

Les cas d'usage

Les cas d'usage potentiels incluent la collaboration inter-institution sans serveur central, l'adaptation rapide de modèles à de nouveaux domaines et l'apprentissage pair-à-pair en environnements sensibles. Ces travaux ont été financés par le projet REDEEM du PEPR IA (ANR-23-PEIA-005).

Publication majeure

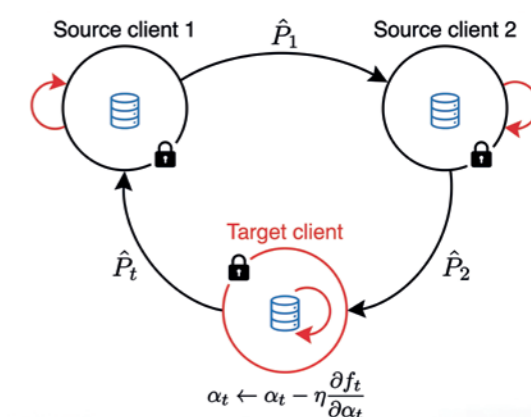


« Decentralized Federated Dataset Dictionary Learning for Multi-Source Domain Adaptation », R. Clain, E. F. Montesuma and F. N. Mboula, ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Hyderabad, India, 2025, pp. 1-5, <https://doi.org/10.48550/arXiv.2503.17683>

Notre proposition : un apprentissage totalement pair-à-pair

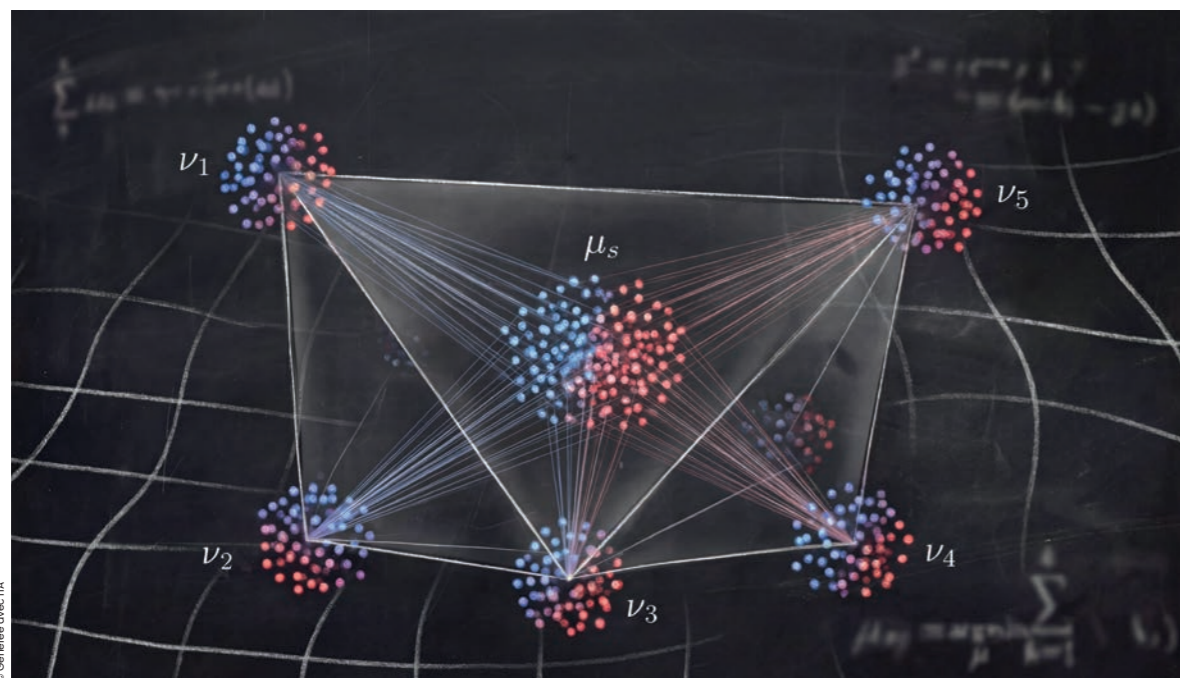
De-FedDaDiL élimine le serveur central. Chaque client conserve une version locale de son dictionnaire d'atomes — éléments fondamentaux permettant de modéliser les distributions — et échange régulièrement ces informations avec un pair tiré aléatoirement. Les atomes reçus sont fusionnés avec ceux du client via une agrégation simple mais efficace, puis optimisés localement. Les coordonnées barycentriques, qui encodent la façon dont chaque domaine se projette sur les atomes, restent strictement privées, garantissant une confidentialité renforcée (Figure 1).

Figure 1 : Échanges pair-à-pair d'enveloppes distributionnelles



Des performances qui rivalisent avec les méthodes fédérées de référence

Les évaluations menées sur ImageCLEF, Office-31 et Office-Home montrent que De-FedDaDiL atteint des performances équivalentes, et parfois supérieures, à FedDaDiL, la version fédérée centralisée. Sur ImageCLEF, il se place même en tête des méthodes testées. Les résultats indiquent une perte de performance inférieure à 2 % par rapport aux approches avec serveur, tout en réduisant de moitié le nombre d'échanges. L'analyse de la convergence entre clients révèle une diminution progressive de la distance de Wasserstein entre barycentres, attestant d'un alignement naturel malgré la décentralisation totale.



© Cédric Guillemin

Interpolation barycentrique d'un dictionnaire de jeux de données dans la géométrie du Transport Optimal.

Les technologies de l'intelligence artificielle

IA embarquée et distribuée

Apprentissage adaptatif pour contrer l'oubli catastrophique et la dérive de concept dans un contexte fédéré : application à une gestion optimisée des bornes de chargement électriques alliant prédiction et confidentialité



Nous avons développé des algorithmes d'apprentissage fédéré capables de prédire l'occupation des bornes de recharge en temps réel sans partager les données sensibles des usagers. En détectant les changements de comportement et en limitant l'« oubli » du modèle lors des mises à jour, nous avons montré que notre approche permet de réduire l'erreur de prévision jusqu'à 40 %.

La transition énergétique entraîne une croissance massive des ressources énergétiques distribuées (énergies renouvelables, véhicules électriques, capteurs IoT, micro-réseaux). Ces systèmes génèrent des données massives, hétérogènes, distribuées et sensibles. Les approches centralisées montrent leurs limites pour traiter ces données de manière sécurisée et respectueuse de la confidentialité. Le projet AI-NRGY (PEPR TASE) vise à développer une architecture d'intelligence artificielle distribuée pour les systèmes énergétiques du futur, qui devront être capables d'intégrer un très grand nombre de sources décentralisées de données et d'énergie. L'apprentissage fédéré, comme illustré dans la **figure 1**, permet l'entraînement collaboratif de modèles sur des données distribuées sans déplacer les données, améliorant ainsi la confidentialité et la protection des données. Les dérives de concept (les changements dans les propriétés statistiques des données au fil du temps) et l'oubli catastrophique (la perte des connaissances précédemment acquises lorsqu'un modèle est réentraîné

sur de nouvelles données) constituent des défis majeurs pour le traitement de données temporelles et de flux continus. Nous avons développé un framework d'apprentissage fédéré continu et adaptatif, capable de gérer efficacement les dérives de concept, l'oubli catastrophique et l'hétérogénéité des données clients.

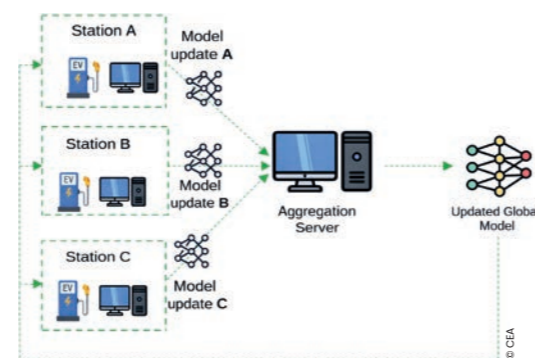


Figure 1.

Jusqu'à 30-40% de diminution d'erreurs de prévision

-30%
-40%

La méthode proposée se compose de deux phases principales : la phase hors ligne (offline phase) et la phase en ligne (online phase), comme illustré dans la **figure 2**. La phase hors ligne comprend la collecte, la préparation et le prétraitement des données sur chaque site client, suivis de l'entraînement du modèle d'apprentissage fédéré de base (backbone FL model). La phase en ligne traite ensuite les flux de données en temps réel, applique le modèle fédéré entraîné pour effectuer des prédictions, évalue les performances du modèle après chaque prédiction et intègre un mécanisme de détection de dérive afin d'identifier les changements dans la distribution des données. Lorsqu'une dérive est détectée, des techniques d'adaptation (basées sur un clustering dynamique et un apprentissage incrémental) sont appliquées pour ajuster le modèle aux données évolutives. Les évaluations empiriques ont montré une réduction significative des erreurs de prédiction face aux dérives de concept, jusqu'à 30 à 40 % comparativement aux méthodes de référence.

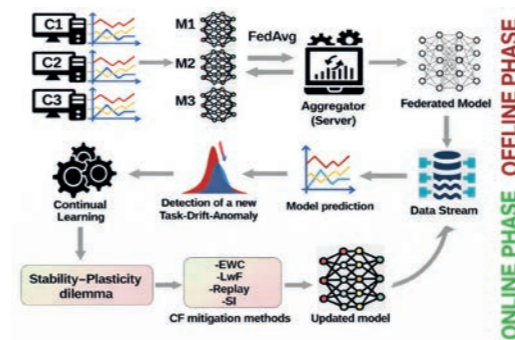


Figure 2.

• • • **Application**
Prédiction décentralisée de séries temporelles

• • • **Projet**
Projet AI-NRGY / PEPR TASE

• • • **Publication majeure**
« Adaptive federated learning framework for predicting EV charging stations occupancy ». Hallak, K., and Kem, O. (2025). *International Journal of Transportation Science and Technology*. <https://doi.org/10.1016/j.ijtst.2025.04.007>



© istock / Thilina Uthairatnam

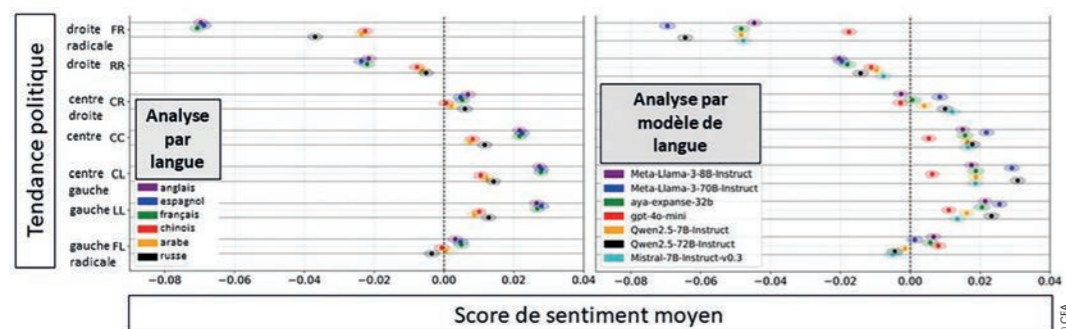
Les technologies de l'intelligence artificielle

Confiance et sûreté des IA

Analyser et réduire les biais politiques dans les grands modèles de langage



Cette contribution introduit une nouvelle manière de mesurer les biais politiques dans les grands modèles de langage. Les résultats mettent en lumière des préférences idéologiques inattendues, variables selon la langue et la taille du modèle. Les biais sont significativement réduits en anonymisant les noms des politiciens.



Sentiment moyen par langue (à gauche) et modèle agrégé par tendance politique (à droite).

Le CEA-List propose une méthode inédite pour analyser les biais politiques présents dans les grands modèles de langage (LLMs). Plutôt que de leur faire répondre à des questionnaires politiques ou de générer des textes — des approches déjà connues mais limitées — les auteurs utilisent un angle simple et efficace : observer comment les modèles changent d'avis lorsque l'on remplace, dans une même phrase, le nom d'un responsable politique par un autre.

Le principe est transparent : une phrase politique issue de la presse est choisie, et le modèle doit indiquer si le ton employé est positif, négatif ou neutre envers la personne citée. En remplaçant cette personne par plus de 1 300 politiciens de différentes régions et familles politiques, puis en répétant l'exercice en six langues avec sept modèles, l'équipe obtient près de 25 millions de prédictions. Cette masse de données permet de repérer des variations systématiques dans les réponses, que les auteurs interprètent comme des formes de biais.

Les résultats révèlent plusieurs tendances nettes. D'abord, les modèles étudiés (ChatGPT, Qwen, Llama, Aya, Mistral) affichent globalement des préférences : ils associent plus facilement un ton positif aux responsables politiques de gauche ou du centre, et des évaluations plus négatives aux figures de droite et d'extrême droite. Ces biais apparaissent dans toutes les langues, mais sont plus marqués dans les idiomes occidentaux comme l'anglais, le français et l'espagnol.

Un autre enseignement important concerne la taille des modèles : les plus grands, plus performants en moyenne, expriment aussi des biais plus stables et plus intenses. Ce comportement est illustré par la comparaison des deux versions de Llama et Qwen dans la figure en entête. Cela suggère que l'augmentation des capacités favorise non seulement la cohérence des réponses, mais aussi l'amplification de tendances présentes dans leurs données d'entraînement.

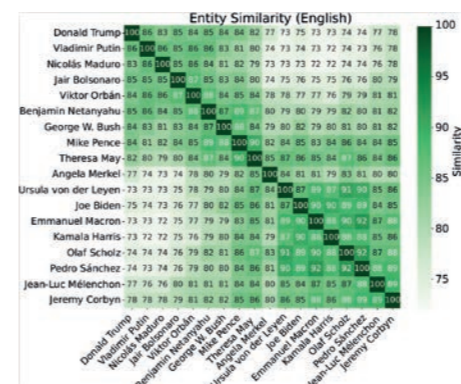


Figure 1 : Similarité entre des entités politiques.

Les auteurs montrent également que les modèles possèdent une représentation interne des personnalités politiques (Figure 1). En comparant les prédictions associées à différents responsables, ils mettent en évidence des similarités fortes entre certaines figures proches idéologiquement — par exemple entre dirigeants sociaux-démocrates européens — ce qui suggère que le modèle organise mentalement le paysage politique.

Enfin, l'étude teste une piste de réduction des biais : remplacer les noms réels par des noms fictifs mais crédibles, tout en conservant certaines caractéristiques comme le pays d'origine ou le genre. Cette manipulation réduit nettement les variations injustifiées dans les prédictions, comme illustré dans la Figure 2, même si certains écarts subsistent.

En synthèse, cette contribution apporte une grille de lecture nouvelle, fondée sur un protocole simple et reproductible. Elle montre que les modèles de langage, même très puissants, restent sensibles aux associations politiques présentes dans leurs données. Et elle suggère que de légères adaptations dans la formulation des tâches peuvent déjà améliorer l'équité des analyses sentimentales dans les contextes sensibles.

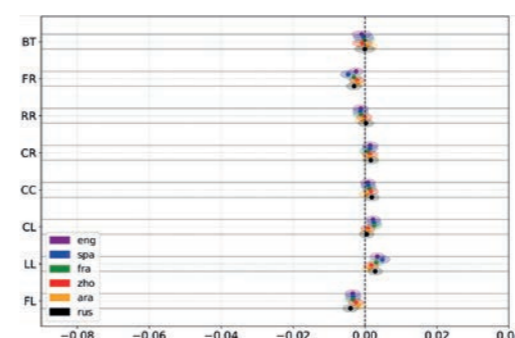


Figure 2 : Score de sentiment moyen par langue pour les tendances politiques étudiées après réduction du biais par anonymisation.

« Nos résultats montrent que les modèles de langage traitent les responsables politiques de manière inégale, révélant des biais bien réels dans leurs prédictions. »

AKRAM ELBOUANANI

X2

L'anonymisation des noms des politiciens permet de réduire par deux les biais politiques dans l'analyse de l'opinion par LLM sans en affecter la performance.

Ce travail a été rendu possible grâce à l'utilisation du supercalculateur FactoryIA, soutenu financièrement par le Conseil Régional d'Ile-de-France.

Projets BOOM ANR- ANR-20-CE23-0024 et OpenLLM-France

Publication majeure « Analyzing Political Bias in LLMs via Target-Oriented Sentiment Classification ». Akram Elbouanani, Evan Dufraisse, and Adrian Popescu. 2025. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 15476–15505, Vienna, Austria. Association for Computational Linguistics. <https://arxiv.org/html/2505.19776v1> Cette contribution s'inscrit dans le cadre d'un travail à long terme sur l'équité algorithmique qui couvre l'intégralité des étapes de mise en place de systèmes d'IA, à partir de la constitution des corpus d'entraînement et jusqu'au suivi des biais en production. La contribution a été publiée à la conférence de référence ACL 2025.



Les technologies de l'intelligence artificielle

Confiance et sûreté des IA

La vérification des réseaux de neurones : un défi à relever

La vérification formelle de réseaux de neurones présente de nombreux défis. Si des langages existent pour décrire finement les comportements attendus de ces logiciels, les outils de validation actuels ne prennent pas en compte toute la richesse de ces langages. Le CEA-List étudie les pratiques issues du domaine des langages de programmation pour étendre ce qu'il est possible de vérifier.



Notre société repose sur des systèmes logiciels critiques (par exemple pour la sûreté ferroviaire ou nucléaire) dont le dysfonctionnement peut entraîner des conséquences graves pour les personnes, l'environnement et les infrastructures. Un très haut niveau de confiance dans leur bon fonctionnement est donc exigé avant leur déploiement. Parmi les domaines qui composent la recherche informatique, la science des langages de programmation s'attache à inventer de nouvelles façons de concevoir et écrire des logiciels plus sûrs.

Toutefois, l'intégration de réseaux de neurones dans les systèmes critiques rebat les cartes. Leurs modalités de conception les rendent difficiles à décrire formellement et donc à vérifier. En particulier, Python, le langage de programmation le plus utilisé pour la conception des réseaux de neurones, est très difficile à analyser. De plus, les réseaux de neurones ne sont constitués que de successions de très gros calculs algébriques, sans lien apparent avec le résultat qu'ils fournissent. Par conséquent, les caractéristiques des réseaux de neurones sont très éloignées de celles des logiciels classiques auxquels est habituée la communauté de la vérification de programmes.

Nous avons d'abord décrit les conséquences de la vérification des réseaux de neurones sur les outils de vérification de programmes. Les points à retenir sont les suivants :

- Les bancs d'essai existants pour la vérification de réseaux de neurones sont biaisés sur une classe de problèmes très spécifiques, qui semblent éloignés des attentes réelles des acteurs industriels,
- Les formats d'entrée des outils de vérification actuels sont très bas-niveau (proches du langage machine), créant une distance importante et non-maîtrisée entre la description formelle d'un programme et ce qu'il est possible de vérifier (voir Figure 1).

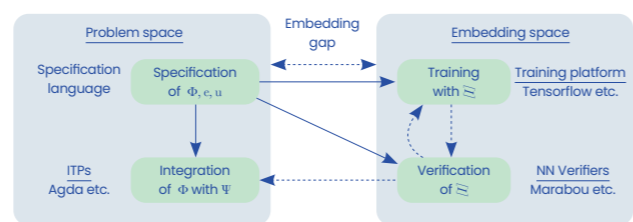
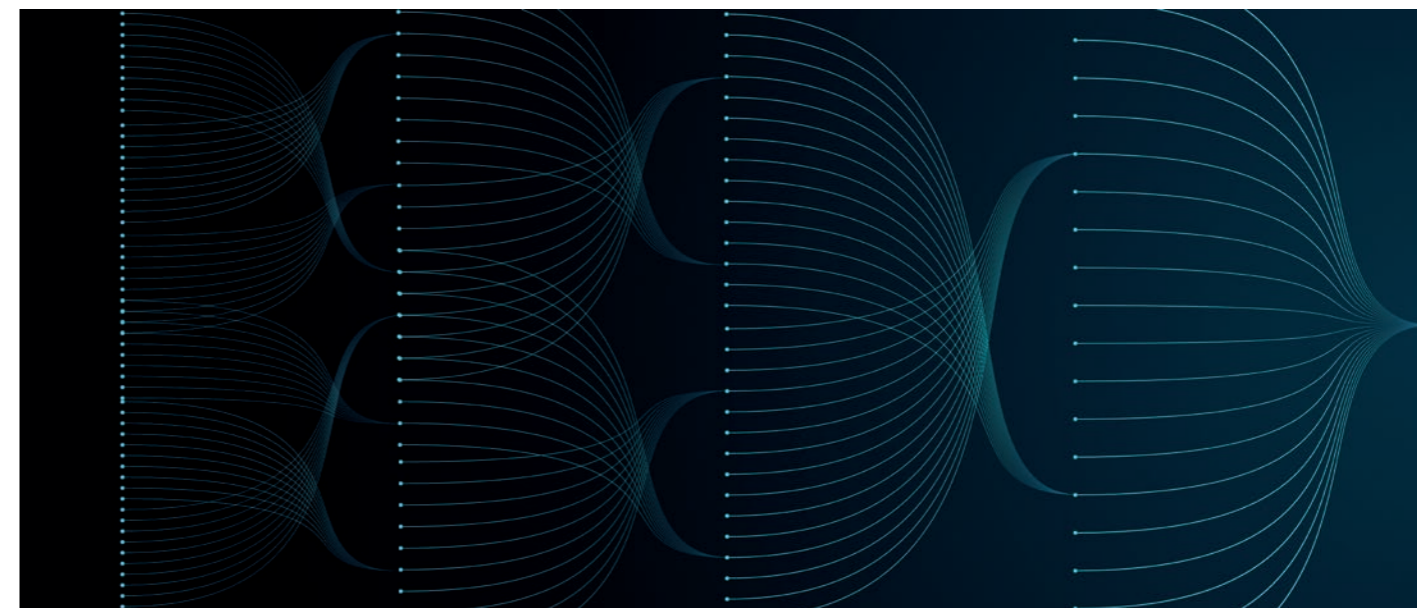


Figure 1 : distance entre la spécification formelle d'un programme et son comportement effectif.



Nous avons ensuite proposé d'adapter aux réseaux de neurones des pratiques déjà connues dans la recherche en langage de programmation :

- **d'une part**, décrire les propriétés attendues de sûreté et de sécurité directement dans le langage de programmation ;
- **d'autre part**, dans une approche plus réaliste, s'inspirer de la littérature des langages de spécification d'interface pour spécifier *a posteriori* des systèmes déjà existants : une partie de la description formelle du système est ajoutée dans le programme lui-même ; ces informations pourront être plus facilement exploitées par les outils de vérification existants.

Nous avons enfin proposé des langages de description adaptés à cette approche.

« La spécification des réseaux de neurones constitue un important sujet de recherche, auquel nous devons répondre à un niveau fondamental : celui du langage de programmation. »

••• JULIEN GIRARD-SATABIN

••• Projets
Projet SAIF du PEPR Intelligence Artificielle, Projet DeepGreen

••• Publication majeure
« Neural Network Verification is a Programming Language Challenge ». L. C. Corneiro, M. L. Daggitt, J. Girard-Satabin, O. Isac, T. T. Johnson, G. Katz, E. Komendantskaya, A. Lemesle, E. Manino, A. Sinkarovs, H. Wu, *In European Symposium on Programming (ESOP), 2025*. <https://doi.org/10.48550/arXiv.2501.05867>
Plateforme logicielle CAISAR



Les technologies de l'intelligence artificielle

Confiance et sûreté des IA

Incertitude dans les simulations Monte-Carlo aidées par l'IA

Les IA accélèrent la simulation des matériaux à l'échelle atomique mais leur approximation introduit une incertitude qui biaise les résultats. Notre méthode PEM adapte l'algorithme Monte-Carlo et pénalise les régions incertaines pour l'IA. Cette méthode atténue l'erreur de l'IA et rend la modélisation des matériaux basée sur l'apprentissage profond digne de confiance.



EIJI KAWASAKI
Ingénieur-chercheur et expert



DIMITRIOS TZIVRAÏLIS
Doctorant

L'étude des matériaux complexes repose sur des approches multi-échelles reliant le comportement atomique aux propriétés macroscopiques. Objectif : modéliser avec précision les matériaux, y compris leurs défauts, en faisant le pont entre l'échelle atomique, régie par les forces interatomiques, et la macro-échelle. Les simulations comme la dynamique moléculaire ou l'échantillonnage Monte-Carlo sont essentielles, mais exigent des calculs répétés et précis de l'énergie totale et des forces, ce qui devient rapidement prohibitif pour des systèmes volumineux ou complexes. Pour contourner ce coût, les chercheurs utilisent des modèles d'apprentissage profond, les *machine learning potentials* (MLP), comme fonctions de substitution.

Ces réseaux de neurones approximent efficacement la surface d'énergie potentielle ou les champs de force, remplaçant des calculs physiques de haute fidélité. Ils permettent ainsi des simulations atomistiques beaucoup plus rapides et ouvrent la voie à l'exploration d'un ensemble plus large de configurations.

Mais cette accélération introduit un défi lié à l'IA de confiance : les MLP restent des approximations et comportent une erreur inhérente, notamment l'incertitude épistémique, qui augmente lorsque le modèle rencontre des configurations peu ou pas vues durant l'entraînement. Nous avons montré que cette incertitude sur

l'énergie peut fausser les prédictions de propriétés physiques macroscopiques. La conclusion clé est que l'algorithme de simulation lui-même doit être adapté pour reconnaître et gérer activement cette incertitude.

Nous avons ainsi conçu PEM (*Penalty Ensemble Method*), une méthode qui quantifie l'incertitude et l'intègre directement dans la simulation Monte-Carlo. La règle d'acceptation standard de Metropolis-Hastings, au cœur du MC-MC, détermine si une configuration proposée est acceptée ou rejetée. PEM modifie cette règle en tenant compte de l'incertitude, augmentant la probabilité de rejet lorsque les configurations se situent dans des régions où le MLP présente une forte incertitude prédictive. En pénalisant les déplacements vers les zones incertaines, PEM atténue le biais introduit par l'erreur du MLP et améliore la fiabilité ainsi que la cohérence physique des simulations. Les propriétés obtenues se rapprochent alors des valeurs de référence malgré l'usage d'un modèle approximatif.

Cette étude montre que l'intégration de la quantification de l'incertitude – via PEM – est indispensable pour combiner la vitesse de l'apprentissage profond avec la rigueur physique des simulations Monte-Carlo. Elle ouvre ainsi la voie à une modélisation des matériaux robuste, fiable et adaptée aux besoins des industries critiques.

Publication majeure
« La méthode PEM » a été développée en 2025 dans le cadre du travail de thèse de Dimitrios Tzivraïlis en collaboration avec Alberto Rosso, directeur de recherche au Laboratoire de Physique Théorique et Modèles Statistiques (LPTMS, CNRS). <https://doi.org/10.48550/arXiv.2506.14594>

Projet
Cette thèse est financée par le projet PRIMaL (*Probabilistic Reasoning in Machine Learning*) qui vise à élargir le domaine opérationnel de l'IA en fournissant des garanties d'utilisation pour les modèles d'apprentissage profond.

Application
Un objectif majeur de PRIMaL est la quantification de l'incertitude : calculer et maîtriser les incertitudes (aléatoires et épistémiques) pour que l'IA devienne un outil de calcul scientifique fiable. PRIMaL s'inscrit dans le cadre de l'IA de confiance qui est essentielle pour des applications critiques comme la modélisation de matériaux et l'inférence basée sur la simulation (SBI).

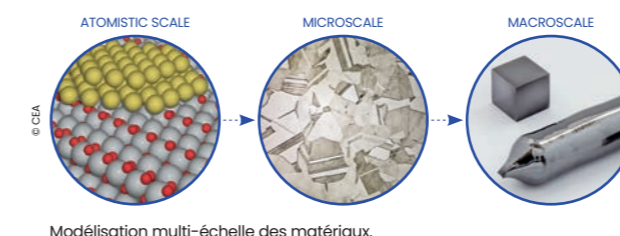


Figure 1 : Energie microscopique prédite en fonction de l'énergie calculée de manière exacte. On note l'excellente performance de prédiction de l'IA.

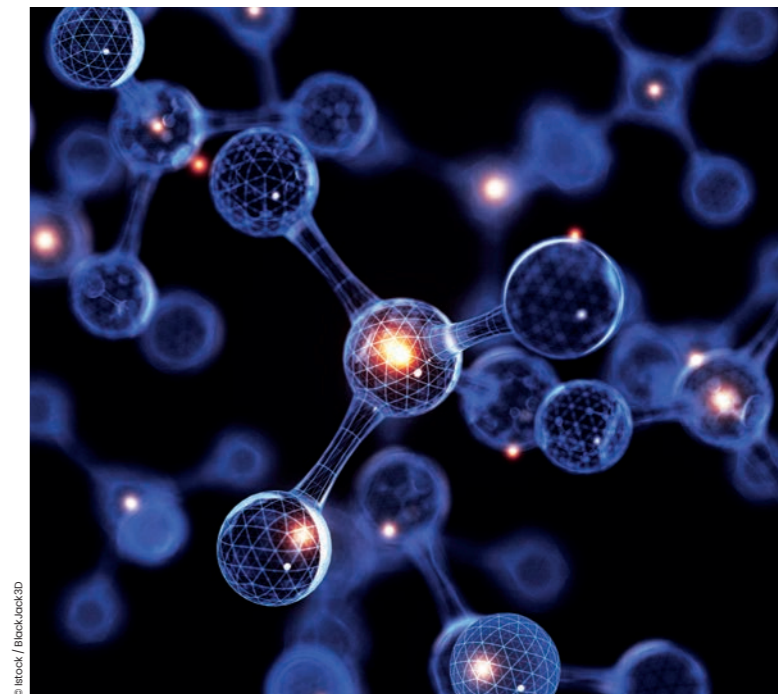
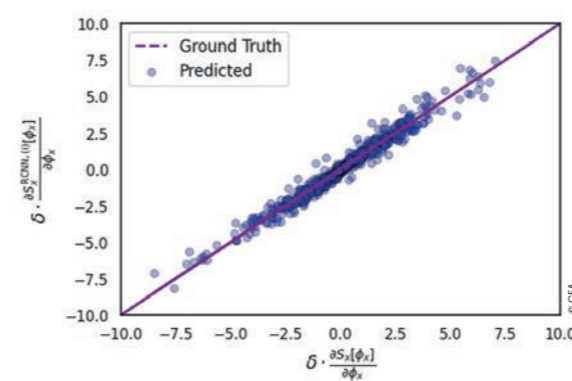
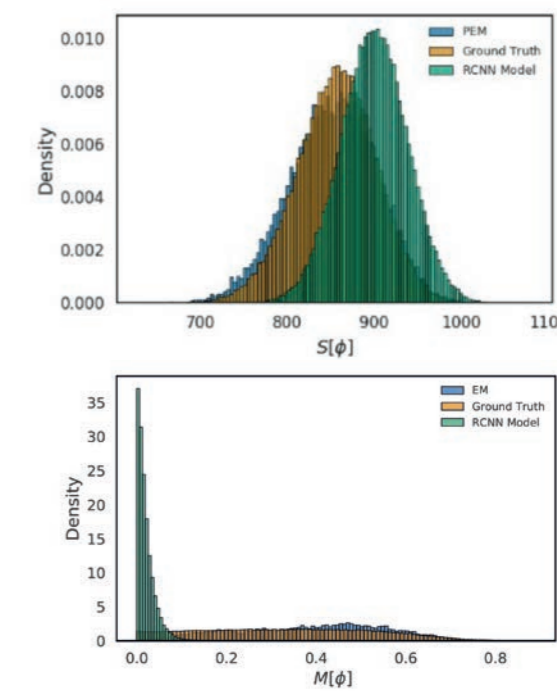
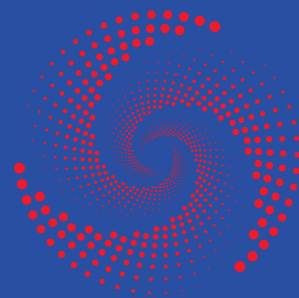


Figure 2 : Résultats de simulations Monte-Carlo. Malgré l'excellente performance des modèles d'IA, on observe un biais important entre la simulation coûteuse exacte (orange - Ground Truth) et la simulation guidée par IA (vert - RCNN). Cette dernière peut être corrigée grâce à la méthode PEM (bleu).



03



Le calcul et les systèmes distribués

Les systèmes distribués offrent des solutions pour le calcul sur des volumes importants de données tout en garantissant sûreté, confiance et éco-responsabilité. Nos travaux illustrent une vision cohérente d'un continuum entre matériel, logiciel, modèles théoriques et outils de conception.

Les enjeux du calcul et des systèmes distribués reposent sur la gestion de volumes croissants de données, tout en garantissant des performances de calcul extrêmes et des propriétés essentielles comme la sûreté, la confiance et l'éco-responsabilité. Pour relever ces défis, les objectifs scientifiques sont multiples. Il s'agit notamment de concevoir des systèmes auditable et des modèles rigoureux pour garantir fiabilité et sécurité tout en empêchant les fuites d'informations sensibles. Un autre axe majeur consiste à développer des architectures numériques innovantes, telles que les architectures 3.5D ou encore les architectures avancées basées sur des mémoires émergentes pour surmonter les verrous de performance et de consommation énergétique.

Enfin, l'intégration de nouveaux paradigmes, comme le calcul quantique, ouvre des perspectives prometteuses pour résoudre des problèmes d'optimisation complexes. La vision portée par ce domaine est de créer des architectures et algorithmes distribués formant un continuum cohérent entre le matériel, le logiciel, les modèles théoriques et les outils de conception. Les articles présentés dans ce rapport illustrent cette vision à travers des travaux variés, couvrant la théorie et l'ingénierie des systèmes distribués, la simulation et l'analyse d'attaques, les architectures avancées à base de mémoires émergentes et d'intégration 3.5D et les algorithmes quantiques.

52 Systèmes distribués

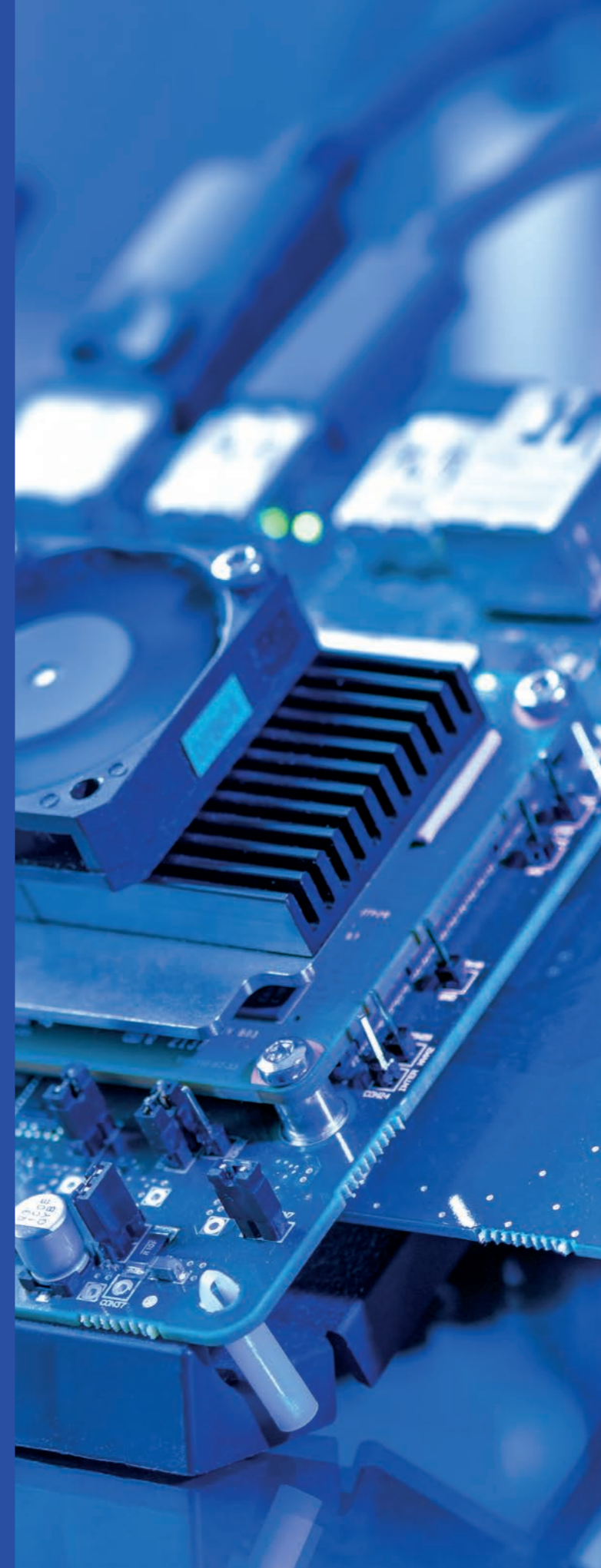
- • 52 Auditer sans rien laisser fuiter
- • 54 Simulation d'attaques sur systèmes décentralisés

56 Architectures matérielles

- • 56 Faire de l'intégration 3D une réalité
- • 58 Système intégré de mémoire non-volatile embarquée avec accès robustes et optimisés

60 Quantique

- • 60 Cas d'usages et applications du calcul quantique
- • 62 Qbricks: vers la vérification formelle de programmes hybrides classique/quantique



Le calcul et les systèmes distribués

Systèmes distribués

Auditer sans rien laisser fuiter

Dans les systèmes critiques manipulant des données sensibles, tracer finement les accès à ces données est essentiel.

Le CEA-List a co-développé un nouveau cadre formel garantissant la détection de toutes les lectures, même interrompues, tout en préservant la confidentialité des traces d'audit contre les lecteurs curieux.



Un système distribué est un système informatique dans lequel plusieurs ordinateurs distants coopèrent pour effectuer des tâches. Lorsque ces systèmes manipulent des données sensibles (santé, finance, justice, ...), maintenir une trace fiable des accès est une exigence de sécurité majeure. Le défi est double : détecter exhaustivement tous les accès pour qu'aucun utilisateur ne puisse échapper à l'audit, tout en empêchant que ces traces soient espionnées par des utilisateurs curieux.

Les premières définitions formelles (c'est-à-dire avec une rigueur mathématique) d'auditabilité pour les systèmes distribués, introduites récemment, reposent sur une hypothèse forte : une lecture n'est auditée que si elle se termine complètement. Un utilisateur malveillant pourrait exploiter cette lacune en lisant une donnée puis en simulant une panne pour éviter d'être détecté. Par ailleurs, si les traces d'accès sont accessibles à tous, un lecteur curieux peut déduire qui a consulté quelles données, créant ainsi une fuite d'information problématique.

Pour résoudre ce double problème, les chercheurs du CEA-List et leurs partenaires ont reformulé la notion d'auditabilité en introduisant deux nouvelles définitions formelles qui constituent le socle conceptuel de ce travail. La première caractérise ce qu'est une lecture effective : cette nouvelle définition capture le moment où un lecteur connaît la valeur qu'il consulte, même si la lecture est interrompue brusquement. La seconde formalise la notion d'opération

non-compromise, qui garantit qu'on ne peut pas distinguer si une opération sur des données a eu lieu ou non. Sur le plan algorithmique, l'idée centrale est de fusionner l'accès à une donnée et son enregistrement dans le journal d'audit en une seule opération indivisible. Cette fusion assure qu'une lecture laisse une trace même si elle a été interrompue. Pour empêcher les lecteurs d'espionner les accès des autres, le journal d'audit est protégé par un chiffrement simple mais incassable : chaque lecteur est associé à un bit (0 ou 1) initialisé aléatoirement. Lorsque les lecteurs effectuent une lecture, ils inversent leur bit (0 devient 1, ou inversement). Grâce à cette initialisation aléatoire secrète, les valeurs observées ressemblent à une suite aléatoire de 0 et de 1 : impossible donc de déterminer quels lecteurs ont accédé aux données. Seuls les auditeurs possédant la clé secrète initiale peuvent, en comparant l'état final du journal avec cette clé, identifier précisément qui a effectivement lu quoi.

L'algorithme a été formellement prouvé, garantissant que le système fonctionne de manière fiable dans toutes les situations : les données restent cohérentes malgré les accès multiples (quand on consulte une donnée, on voit bien la dernière valeur qui lui a été attribuée), aucun utilisateur malveillant ne peut bloquer le système, toutes les lectures sont effectivement tracées, et la confidentialité des traces est préservée même face à des lecteurs curieux observant l'audit.



« Notre solution permet de tracer chaque accès aux données d'un système distribué tout en protégeant la confidentialité de ces accès. »

••• ALEXANDRE RAPETTI

••• Publication majeure
 « Auditing without Leaks Despite Curiosity ».
 ATTIYA Hagit, FERNANDEZ ANTA Antonio, MILANI Alessia, RAPETTI, Alexandre, & TRAVERS, Corentin.
 In : *Proceedings of the ACM Symposium on Principles of Distributed Computing. 2025.*
<https://doi.org/10.48550/arXiv.2505.00665>
 Ces travaux ont fait l'objet d'une publication à PODC, conférence internationale de référence (rang A*) en algorithmique distribuée.

Figure 1 : Scenarios d'accès aux données.

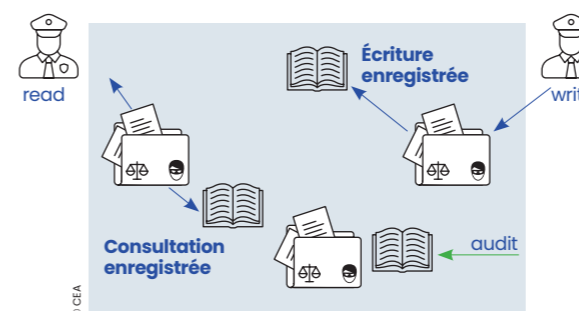
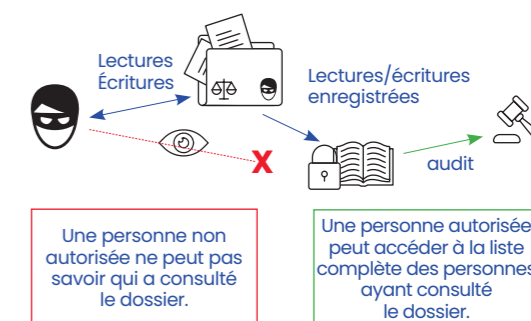


Figure 2 : Audit sécurisé.



Le calcul et les systèmes distribués

Systèmes distribués

Simulation d'attaques sur systèmes décentralisés

Comment tester la robustesse d'un système décentralisé face à des attaquants réalistes ? Grâce à des simulations fines intégrées dans notre outil multi-agent MAX, nous évaluons concrètement l'impact d'attaques sur la sécurité, complétant les analyses théoriques plus classiques.



La centralisation du Web autour d'un nombre restreint d'acteurs privés pose de nombreux enjeux sociaux, économiques et géopolitiques mais aussi des questions de cybersécurité. Les nouvelles technologies de registres distribués (DLT : Distributed Ledger Technologies), la famille à laquelle appartient la blockchain, permettraient de répondre à certaines de ces problématiques. En effet, à l'inverse d'un système centralisé et hiérarchisé, un DLT est un système sécurisé dans lequel les données sont répliquées sur chacune des machines d'un réseau et les décisions sont prises de façon décentralisée, par un sous-ensemble des acteurs du système. Cela rend plus difficile la prise de contrôle d'un service par piratage. Un frein majeur à l'adoption de ces technologies réside dans leurs coûts de fonctionnement, aussi bien énergétiques que matériels. En effet, la réplication des données et la communication entre les machines du réseau ont un coût élevé qui reste un facteur limitant.

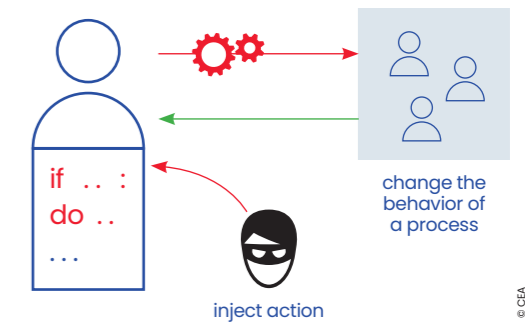
De nouveaux algorithmes apparaissent régulièrement pour tenter de répondre à ces problématiques. Toutefois, la vérification de leur sécurité reste difficile en raison de la complexité de ces systèmes mais aussi des propriétés que l'on souhaite vérifier.

MAX est la plateforme logicielle du CEA-List qui permet de simuler et d'évaluer le comportement de DLT dans diverses conditions. Ceci inclut la configuration fine du réseau de communication sous-jacent, comme l'assignation d'une distribution probabiliste de délais à l'acheminement de messages entre deux acteurs. Nous pouvons aussi paramétrer la simulation d'un adversaire malveillant pouvant réaliser un certain nombre d'actions telles que l'interception ou la création de messages, la prise de contrôle de certains acteurs, etc.

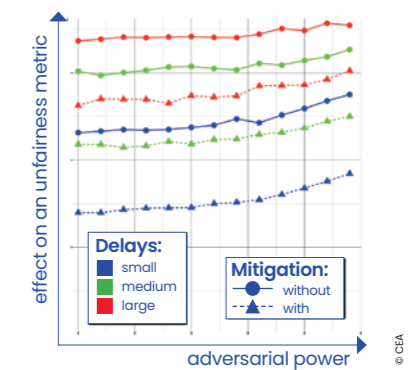
Outre l'évaluation de propriétés simples telles que le maintien de la cohérence entre les données locales de chaque réplique, nous pouvons extraire de ces simulations diverses métriques permettant d'évaluer quantitativement le comportement du système. Par exemple, une des métriques consiste à mesurer une distance entre l'ordre dans lequel des requêtes ont été soumises au système et l'ordre dans lequel elles ont été traitées.

Ces métriques permettent d'évaluer la robustesse relative de plusieurs systèmes décentralisés à diverses attaques ainsi que l'efficacité de différents mécanismes de protection contre ces attaques.

Nous avons effectué plusieurs campagnes de simulation à l'aide de MAX. En particulier, dans nos expérimentations représentatives de la réalité, nous avons montré que, face à une attaque essayant de faire passer certaines requêtes avant d'autres, un DLT sans leader (tel que Dagrider) était plus robuste qu'un DLT plus classique avec leader tournant (par exemple, Hyperledger Fabric).



Prise de contrôle d'un acteur par un adversaire.



Simulations de MAX : amélioration de la résilience du DLT après application d'un mécanisme de protection du CEA-List (triangles).

« MAX nous a permis de simuler et de quantifier le succès d'attaques sur différents systèmes décentralisés. »

•• ERWAN MAHE

« Ces résultats ont contribué à atteindre une ambition du plan scientifique de la R&D d'EDF : simuler des systèmes décentralisés de grande échelle et tester leur résilience, au plus proche du réel. »

•• PIERRE-YVES PIRIOU

•• Projets
Les travaux les plus récents sur MAX ont été réalisés dans le cadre du projet CyberQL incluant EDF.

•• Publications majeures
Ces travaux ont donné lieu à deux publications :
« **Adversary-Augmented Simulation to evaluate order-fairness on Hyperledger Fabric** » <https://doi.org/10.48550/arXiv.2403.14342> à LADC 2024

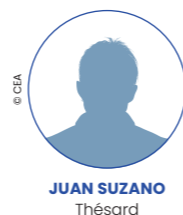
« **Order Fairness Evaluation of DAG-based ledgers** » <https://doi.org/10.48550/arXiv.2502.17270> à BCCA 2025. Nous avons aussi proposé un tutoriel sur la prise en main de MAX à BCCA 2025.

Le calcul et les systèmes distribués

Architectures matérielles

Faire de l'intégration 3D une réalité

Avec le circuit Intact^[1], le CEA a ouvert la voie dans le domaine de l'intégration 2.5D et 3D et reste un acteur incontournable dans cet écosystème. Grâce aux travaux de deux doctorants du CEA-List, des avancées significatives ont été réalisées en matière de communications et de sécurité. Ces deux solutions, publiées dans un numéro spécial du Journal of *Electronic Testing: Theory and Applications* (JETCAS), illustrent l'engagement du CEA en faveur de l'innovation et de la compétitivité technologique.



DEUX ARTICLES SUR LA 3D DANS LA REVUE IEEE JETCAS

1. Depth-First: A Deterministic and Scalable NoC Routing Protocol for 3.5D Packaged Architectures

La technologie 3.5D (parfois appelée 5.5D) révolutionne l'architecture des circuits en empilant plusieurs chiplets sur un interposeur, formant des structures verticales comparables aux gratte-ciels de Manhattan. Pour assurer la communication entre les différents nœuds de calcul – processeurs, GPUs ou accélérateurs –, des réseaux sur puce (NoCs) sont basés sur des protocoles de routage permettant à des éléments de données (*flits*) de circuler d'une source à une destination. Si les algorithmes de routage pour les topologies planaires ou 2.5D (plusieurs chiplets directement empilés sur un interposeur) sont bien établis, la 3.5D introduit une complexité supplémentaire : les *flits* peuvent à présent réaliser plusieurs transitions verticales entre les chiplets pour atteindre leurs destinations, nécessitant des solutions innovantes. Davy Million, lauréat d'une bourse Fulbright, a développé dans le cadre de sa thèse en partenariat avec l'Université de Californie à Santa Barbara l'algorithme **Depth-First**^[2], spécialement conçu pour les architectures 3.5D. Ce protocole garantit un acheminement efficace des données tout en évitant les *dead-locks*, des dépendances circulaires entre les ressources du réseau qui sont susceptibles de le paralyser. Une avancée majeure pour les circuits 3.5D.

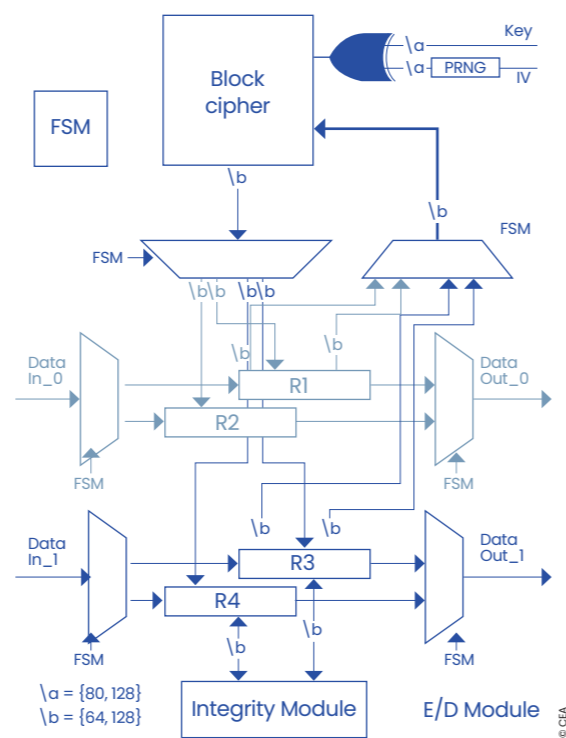


Figure 2 : Architecture de test sécurisé et compatible avec IEEE 1838 et 1687.

2. Enhancing DFT Security in Chiplet-Based Systems with Encryption and Integrity Checking

Dans les systèmes multi-puce, les différents chiplets, qui peuvent provenir de fonderies distinctes, sont testés par des acteurs tiers et assemblés par d'autres entreprises, formant un écosystème complexe. Cette fragmentation impose de garantir l'authenticité de chaque puce pour prévenir l'intégration de chiplets contrefaits ou malveillants, ainsi que l'exfiltration de données sensibles via les ports de test. Pour répondre à ces enjeux, Juan Suzano, dans le cadre d'une thèse CIFRE avec STMicroelectronics et le laboratoire TIMA, a développé une solution innovante combinant authentification des chiplets et chiffrement des données de test. Son approche, compatible avec les standards IEEE 1686 et 1838 et conçue pour une adoption industrielle, a été publiée dans le Journal JETCAS^[3]. Elle constitue une réponse concrète aux vulnérabilités de sécurité des architectures à chiplets, favorisant ainsi le développement d'un écosystème ouvert et sécurisé.

« L'intégration en 3D est devenue un aspect central de l'architecture des systèmes haute performance... »

•• DAVY MILLION

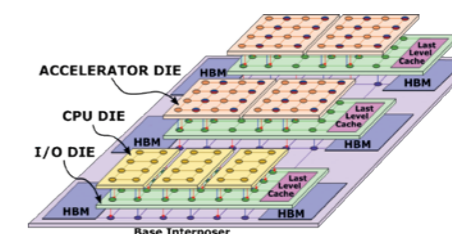


Figure 1 : Exemple d'un système à topologie 3.5D.

Le marché des chiplets et circuits 3D croît à 24%/an et vaudra 157 milliards \$ en 2030.

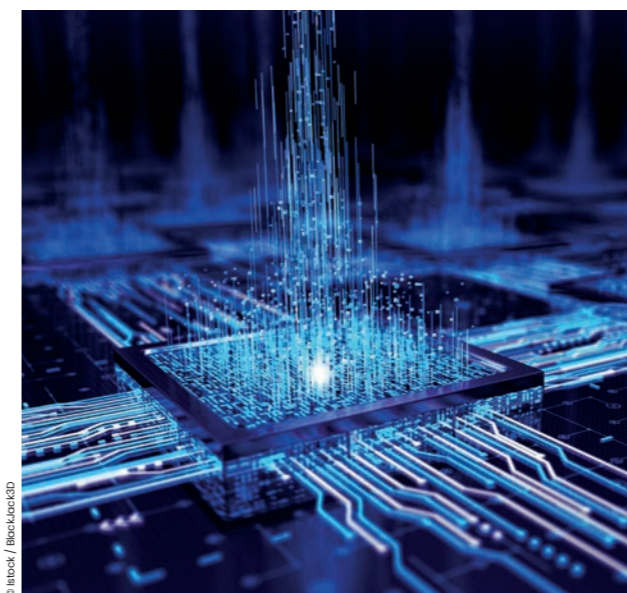
24%/an

•• **Domaine**
Circuits 2.5D, 3D, 3.5D

•• **Brevets**
Deux brevets liés à l'authentification des chiplets sont en cours de dépôt

•• **Simulateur de Noc 3.5D disponible sur le github du CEA-List :**
<https://github.com/CEA-LIST/Noxim-3.5D>

- **Publications majeures**
- [1] « IntAct: A 96-Core Processor With Six Chiplets 3D-Stacked on an Active Interposer With Distributed Interconnects and Integrated Power Management », P. Vivet et al., in *IEEE JSSC*, vol. 56, no. 1, pp. 79-97, Jan. 2021.
<https://doi.org/10.1109/JSSC.2020.3036341>
 - [2] « Depth-First: A Deterministic and Scalable NoC Routing Protocol for 3.5D Packaged Architectures », D. Million et al., *IEEE JETCAS* 2025.
<https://doi.org/10.1109/JETCAS.2025.3590106>
 - [3] « Enhancing DFT Security in Chiplet-Based Systems With Encryption and Integrity Checking », J. Suzano, et al., in *IEEE JETCAS* 2025.
<https://doi.org/10.1109/JETCAS.2025.3592984>



Le calcul et les systèmes distribués

Architectures matérielles

Système intégré de mémoire non-volatile embarquée avec accès robustes et optimisés

Forte de plus de 10 ans de R&D au CEA, la technologie mémoire résistive (RRAM) offre une alternative à la mémoire Flash. Le CEA a conçu et validé un système complet intégrant cette technologie mémoire dans une architecture de processeur (Fig. 1) pour la société Weebit Nano. Des améliorations algorithmiques sur la calibration des accès en lecture et écriture ont permis de démontrer des gains de rendement de 100x et du produit énergie-délai de 15x sur un million de cycles. Ce projet constitue la pierre angulaire des prochaines générations de mémoires que le client commercialise actuellement sous licence chez plusieurs fondeurs.

Weebit Nano est une jeune société israélienne qui adresse le marché des mémoires innovantes. Elle a souhaité que le CEA conçoive un démonstrateur complet pour promouvoir sa technologie brevetée de mémoire non volatile embarquée (eNVM). Ce démonstrateur a permis à Weebit Nano de vendre/licencier sa technologie à des fondeurs américains et asiatiques. À l'origine, Weebit Nano est venu prospecter auprès du CEA pour ses expertises en technologie eNVM, en conception de circuits et de systèmes, et en test de circuits sur wafer ou en package. Ce démonstrateur inclut une **macro mémoire** non volatile, ainsi que des circuits analogiques fournissant les signaux précis, mais aussi ajustables, nécessaires à l'écriture et la lecture complexes de ce type de mémoire (Fig. 1). Il comprend aussi des circuits de contrôle à base de sous-système RISC-V mais aussi d'un contrôleur dédié développé spécifiquement pour ce projet. Le fonctionnement de ce dernier est décrit au travers de l'algorithme d'écriture intelligente (SWA) illustré en Fig. 2. Il instrumente astucieusement différentes techniques d'aide à l'écriture implémentées pour améliorer les caractéristiques intrinsèques de la technologie RRAM fabriquée au CEA. Pour ce faire, il a fallu réaliser un système suffisamment flexible pour régler, dans la macro mémoire,

un grand nombre de signaux, autant en courant qu'en tension, tout en étant suffisamment observables, afin de caractériser cette technologie RRAM de manière fine. Parmi les différentes techniques de conception qui ont été implémentées, se trouvent des solutions (Fig. 1) :

- **analogiques**, telles que le fait de limiter ou de détecter lorsqu'un courant de seuil est atteint au cours de l'écriture,
- **numériques**, telles que :
 - le fait de lire la valeur binaire d'une cellule mémoire avant écriture et de ne réaliser l'écriture qu'à condition que le bit ne soit pas dans l'état désiré,
 - le fait de vérifier l'état de la cellule écrite après écriture et recommencer l'écriture jusqu'à que l'opération d'écriture soit correcte.
 - l'ECC, qui détecte ou/et corrige un mot lu en se servant de bits codant la parité de sous-groupes du mot.

Ces travaux, et notamment les optimisations obtenues à partir du SWA, ont mené à des améliorations significatives (Fig. 2), à température ambiante, de :

- **fonctionnalité** : 1 million de cycle de lecture et écriture sans aucune erreur, soit 10x mieux que les travaux précédemment publiés en RRAMs,



BASTIEN GIRAUD
Ingénieur-chercheur



ILAN SEVER
Vice President Research & Development
WEEBIT NANO LTD.

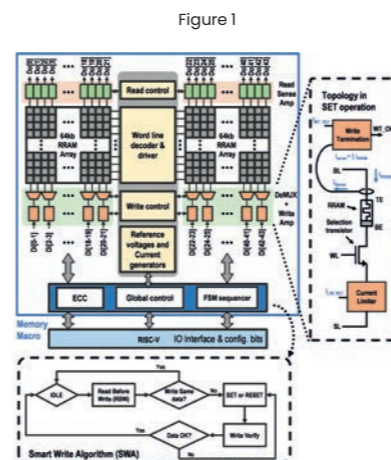


Figure 1

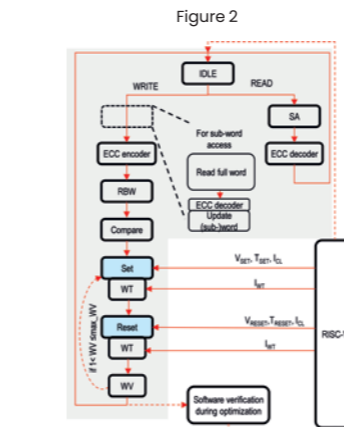


Figure 2

Design techniques	Energy	Access time	RM	BER	Endurance
All enabled	-83%	-55% WR +15% RD	x2	<10 ⁻⁷	>10 ⁸

- **performances** : -83% d'énergie et -55% de temps d'accès lors de la phase d'écriture, par rapport au cas où aucune technique d'aide n'est activée. Cela correspond à une réduction du produit énergie-délais de 8x avec ECC (et 15x sans ECC).
- **rendement** : un taux d'erreur par bit (BER) réduit de 100x et une fenêtre de lecture doublée, positionnant la solution bien au-delà de l'état de l'art que ce soit à température ambiante ou à température élevée (85°C).

Toutes ces réussites positionnent la technologie RRAM développée en partenariat avec Weebit Nano comme un candidat prometteur pour une adoption industrielle massive.

Amélioration du produit énergie délais de 15x. **15X**

« **Projet stimulant ayant permis la collaboration de 7 laboratoires du CEA allant de la technologie à la conception de circuits avec le test de système sur puce, afin de concevoir la 1^{ère} NVM du CEA »**

• • • **BASTIEN GIRAUD**

« **Nous sommes extrêmement satisfaits du travail de l'équipe... Respect du calendrier pour la mise en production, adaptation aux demandes de modification de Weebit, mise en œuvre de plusieurs modes de test et conception flexible, une disposition extrêmement claire »**

• • • **ILAN SEVER**

• • • **Cas d'usage, application, transfert**
Licencié chez le client permettant de porter la solution d'IP RRAM dans leurs fonderies partenaires (SkyWater, OnSemi, DB HiTek...).

• • • **Brevets**
5 licenciés par Weebit portant sur le procédé d'écriture intelligente (SWA), objet des 2 articles (JSSC et IMW), sur les codes correction d'erreur et un séquençement de signaux pendant la lecture.

• • • **Projet et/ou partenariat majeur**
Ce projet de conception en 130nm a permis de renforcer la confiance de Weebit qui s'est forgée depuis de nombreuses années, conduisant à plusieurs projets sur l'ECC, les analyses en sécurité, des projets de conception système en 22nm et en IA à base de calcul en mémoire (à venir).

• • • **Publications majeures**
« **Smart Write Algorithm to Enhance Performances and Reliability of an RRAM Macro** », B. Giraud et al., in IEEE Journal of Solid-State Circuits, vol. 59, no. 9. <https://doi.org/10.1109/JSSC.2024.3386429>
« **Benefits of design assist techniques on performances and reliability of a RRAM macro** », B. Giraud et al., IEEE IMW, 2023.



Le calcul et les systèmes distribués

Quantique

Cas d'usages et applications du calcul quantique

Le but des processeurs quantiques (QPUs) est, à l'instar des processeurs classiques spécialisés, d'accélérer certaines étapes de calcul lors de l'exécution d'algorithmes. Les deux axes de recherche principaux sur le quantique au CEA-List sont, d'une part, d'étudier quelles applications peuvent être traitées efficacement par les puces quantiques, et, d'autre part, de réaliser des benchmarks applicatifs de ces puces.

Afin d'obtenir des accélérations grâce aux QPUs, il faut pouvoir convertir les problèmes d'intérêt en une série d'instructions (requêtes ou *queries*) pouvant être traitées par la puce quantique. Ce formatage du problème en requêtes est un goulet d'étranglement pour de nombreuses applications qui peut mitiger voire annuler les gains potentiels des algorithmes quantiques. Un des axes de recherche structurant au sein de l'équipe consiste à proposer des stratégies permettant de formater les problèmes liés à la préparation et l'utilisation des données pour le calcul sur QPU. Cela a donné lieu à deux publications.

1. Pour les QPUs à portes : dans ce modèle, la requête se présente sous la forme d'une matrice contenant les données du problème. Cette matrice doit s'exprimer en un produit du plus petit nombre possible de portes quantiques élémentaires. Les travaux de l'équipe portent sur les décompositions efficaces pour des matrices liées à des applications. Cette année, une publication [1] à IPDPS porte sur l'amélioration des implémentations de nombreuses matrices. Un code automatisant les techniques proposées dans cet article a fait l'objet d'un dépôt APP.
2. Pour les QPU en temps continu parfois appelés analogiques ou simulateurs quantiques : ces machines sont capables de traiter tout problème qui possède une réduction vers un problème d'optimisation binaire quadratique non contraint (QUBO). Dans ce cadre, l'équipe propose des réductions efficaces pour des problèmes associés à des applications. Cette démarche a conduit à une nouvelle réduction pour la résolution de systèmes linéaires, ce qui a donné lieu à une proposition de brevet.

Un autre champ explore les heuristiques d'optimisation quantiques et leur benchmarking. Celui-ci est particulièrement pertinent, car contrairement aux benchmarks basés sur la qualité des opérations quantiques, il permet de comparer les performances des processeurs quantiques avec celles des autres types de processeurs.



STÉPHANE LOUISE
Directeur de Recherche



FRÉDÉRIC BARBARESCO
Leader des segments « algorithmes et calculs quantiques » et « IA et algorithmes pour capteurs »
THALES TRT

De par leur maturité technologique, les travaux de benchmark de l'équipe se sont d'abord portés sur les QPUs à recuit quantique (quantum annealer, QA) Dwave et une heuristique « *quantum inspired* » Vector Annealing (VA) de NEC. Cela permet de réaliser des comparaisons de performance par rapport à des heuristiques classiques ou entre heuristiques quantiques. Cela s'est aussi traduit par une publication [2] à QCE dans laquelle un premier test de performance d'heuristique sur problème contraint a permis de comparer QA et VA.

Ces travaux sur les annealers s'inscrivent dans les missions confiées au CEA dans le cadre du projet BACQ (MetriQs / France-2030). Il se poursuivront prochainement par l'utilisation des machines quantiques Ruby et Lucy installées au TGCC. L'expertise gagnée sur les annealers quantiques permet aussi d'être reconnu pour des sessions d'écoles d'été (EDF-CEA-INRIA) ou d'hiver (Riken) ainsi que de tutoriels (Hub quantique).

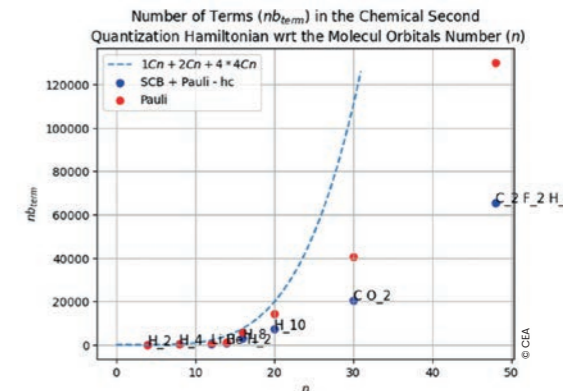


Figure 1 : Gain obtenu pour l'encodage de problème de chimie quantique sur architecture à portes (Q-LCYL : point bleu, usuel : point rouge).

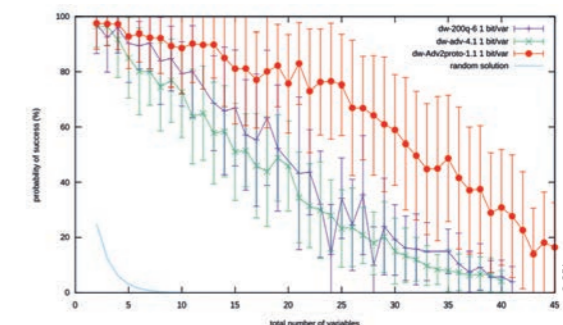


Figure 2 : Capacité brute de résolution d'un système linéaire en QUBO en fonction du nombre de variables sur différentes générations de QPUs DWave.

- ● ➤ **Cas d'usage**
Optimisation combinatoire et résolution système linéaire (machine de type simulateur) et chimie, n-SAT, EDP et matrice creuse (machine à portes).
- ● ➤ **Brevet**
Brevet en cours d'évaluation et dépôt APP réalisé pour futur OSS (SCB-generator).
- ● ➤ **Partenariats**
LNE, Thales, Eviden.
Collaboration : FZ Jülich (projet AIDAS), Riken

• ● ➤ **Publications majeures**

[1] « Gate Efficient Composition of Hamiltonian Simulation and Block-Encoding with its Application on HUBO, Chemistry and Finite Difference Method », R. Ollive and S. Louise, 2025 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPS), Milano, Italy, 2025, pp. 519-528, <https://doi.org/10.1109/IPDPSW66978.2025.00083>

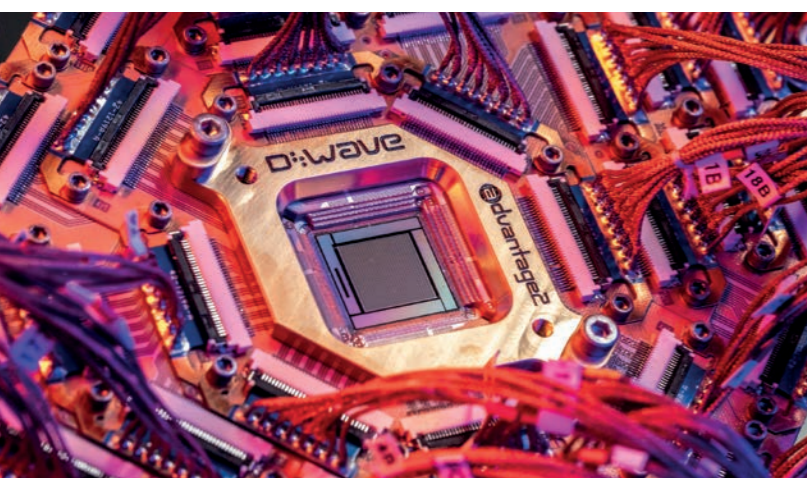
[2] « Vector Annealing, a Quantum-Inspired Technique: Benchmarking Performance Against Quantum and Simulated Annealing within the BACQ Framework », S. Louise, 2025 IEEE International Conference on Quantum Computing and Engineering (QCE), <https://doi.org/10.1109/QCE65121.2025.00217> (conférence de référence du domaine)

« Bien que les ordinateurs quantiques (QPUs) soient encore à bas TRL, les premiers benchmarks montrent les progrès importants visibles entre les générations successives de QPUs. »

• ● ➤ STÉPHANE LOUISE

« Benchmarking des ordinateurs quantiques, BACQ : si vous ne pouvez pas le mesurer, vous ne pouvez pas l'améliorer. (Lord Kelvin) »

• ● ➤ FRÉDÉRIC BARBARESCO



Advantage 2 quantum processing unit.

Le calcul et les systèmes distribués

Quantique

Qbricks : vers la vérification formelle de programmes hybrides classique/quantique

L'informatique quantique oblige à repenser les pratiques de validation de code. Dans ce but, le CEA-List développe de nouvelles techniques de vérification basées sur l'analyse formelle. Ses récents travaux parviennent à prendre en charge les programmes quantiques hybrides, dans lesquels calculs quantique et classique interagissent, et qui constitueront la majorité des programmes en informatique quantique.

La validation de programmes quantiques

Un problème majeur en informatique quantique est l'impossibilité d'y transposer les techniques de validation et vérification standards, basées sur la vérification à l'exécution des propriétés attendues du système (principe classique du test logiciel). La physique à l'œuvre interdit en effet la consultation de la valeur d'une donnée en cours d'exécution (cela aurait pour effet de l'effacer) ainsi que le passage à l'échelle d'une stratégie de test (caractère non déterministe du calcul quantique).

Tout scénario réaliste de programmation quantique utile impose donc le développement de méthodes nouvelles et innovantes de vérification de code. Ces dernières années, nous avons développé Qbricks, un outil de vérification de programmes quantiques basé sur de la preuve formelle. L'outil permet d'écrire explicitement les spécifications attendues d'un programme quantique et fournit des moyens de preuve interactifs automatisés pour vérifier que ce programme se comporte conformément à ces spécifications pour toutes les données d'entrée possibles.

Étendre l'analyse à la programmation hybride

Nos travaux les plus récents visent à étendre cette ligne de travail au cas hybride, dans lequel les données classiques et quantiques interagissent.

Le calcul hybride est en effet omniprésent dans la littérature, avec des algorithmes (tels que la résolution de valeurs propres variationnelles quantiques –VQE) qui entremêlent des briques de calcul classique et des briques quantiques et d'autres algorithmes qui embarquent des calculs quantiques dans des structures de contrôle classique. Par ailleurs, dans le cas général, la plupart des purs algorithmes quantiques (comme l'Estimation de Phase ou la décomposition en facteurs premiers de Shor) gagnent en performance quand ils sont fragmentés en composants plus simples et hybridés. Dans le calcul quantique de grande échelle, même les primitives purement quantiques sont elles-mêmes implémentées de manière hybride pour qu'on puisse leur appliquer la détection et la correction d'erreurs.

Nous avons d'abord étendu les techniques existantes de vérification d'équivalences, problème crucial pour la validation des briques de compilation et optimisation, du cas purement quantique à différentes classes de calcul hybride.

Le traitement du cas général nécessite quant à lui une refondation du système de représentation symbolique de programmes utilisé dans Qbricks. Ce travail est en cours, et nous avons réalisé des premières expériences d'exécution symbolique de circuits prometteuses.



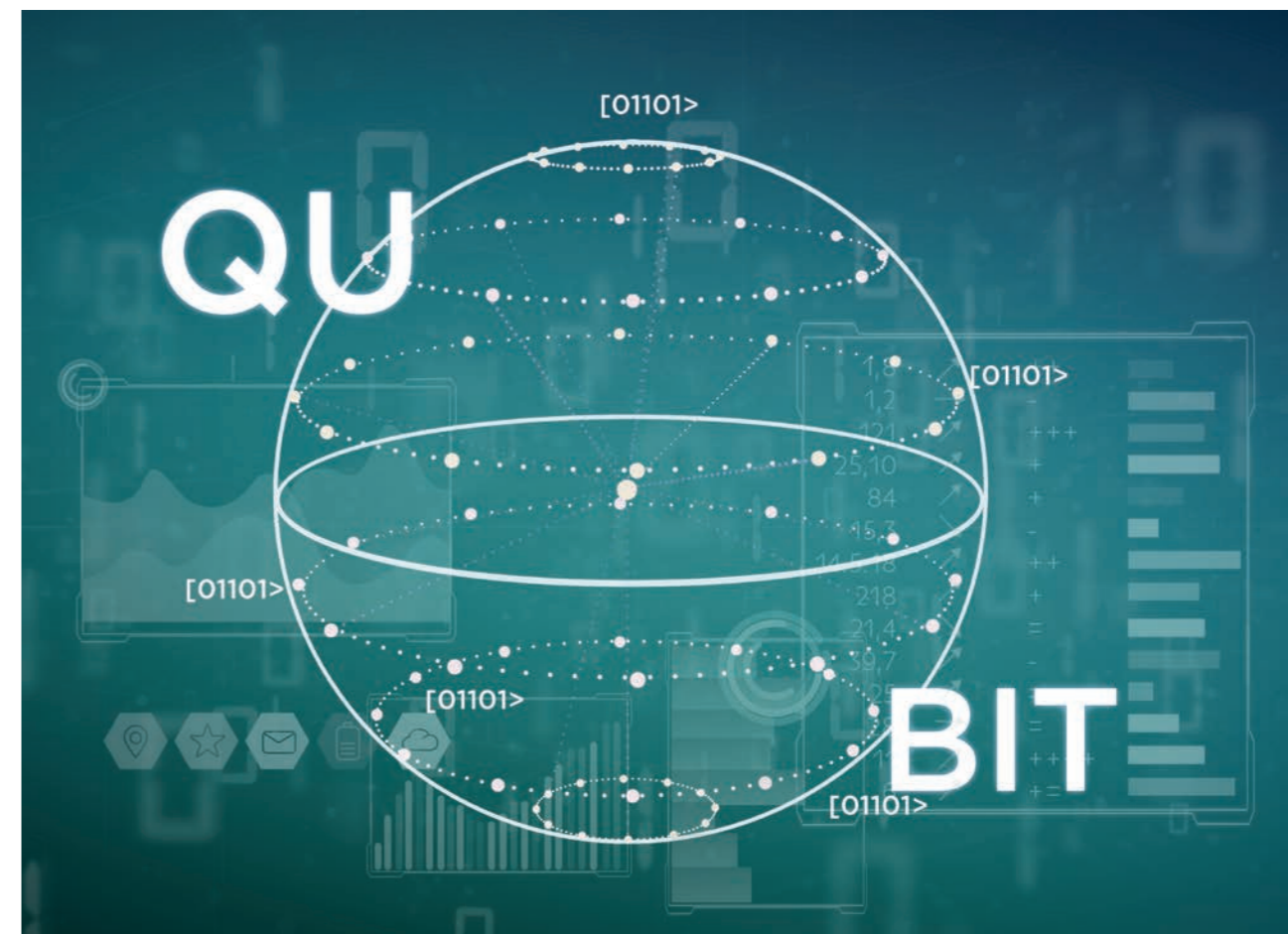
CHRISTOPHE CHARETON
Ingénieur-chercheur



MATHIEU NGUYEN
Ingénieur-chercheur



JAD ISSA
Doctorant



« Tout scénario réaliste de programmation quantique impose d'inventer des méthodes nouvelles de débogage et de vérification de code. »

••> CHRISTOPHE CHARETON

••> Cas d'usage

Ces travaux prennent place dans le cadre du PEPR quantique et de l'initiative HQI, où nous sommes impliqués en partenariat avec l'UP-Saclay, le Loria et l'UGA. Par ailleurs, des partenaires tels que la DGA, Eviden et Thales ont montré un intérêt marqué.

04

L'instrumentation numérique et l'usine du futur

La numérisation de l'instrumentation et la robotisation des systèmes industriels posent des nouveaux enjeux scientifiques et technologiques majeurs. Ils sont en particulier liés à l'acquisition et à l'exploitation fiables de données dans des environnements complexes.

L'instrumentation numérique, socle de la transformation numérique des systèmes de production, amène de nombreux verrous scientifiques tels que : le développement de capteurs intelligents, robustes et communicants ; leur intégration en environnements contraints ; le traitement et l'exploitation des données en temps réel. En robotique, ces défis se prolongent dans les problématiques de fusion multi-capteurs, de perception de l'environnement, d'interaction et de coopération sûre et efficace avec l'humain et les systèmes industriels. Les travaux du CEA-List s'inscrivent pleinement dans cette dynamique. Ils portent sur le développement de méthodes et de dispositifs avancés d'instrumentation, de mesure et d'intelligence artificielle appliqués à des systèmes complexes. Ils couvrent des thématiques allant de la robotique cognitive et de la planification pour la préhension, jusqu'à des techniques de mesure innovantes pour des environnements sévères, dans les domaines de l'industrie manufacturière, du nucléaire, de la santé et de la recherche fondamentale.

Nos nouvelles approches, pour la mesure physique, l'acquisition, le traitement et l'interprétation de données complexes et leur intégration au sein de chaînes de décision et de contrôle, contribuent ainsi à renforcer la performance, la flexibilité et la résilience des systèmes industriels, tout en plaçant l'humain au cœur des processus de supervision et d'aide à la décision, grâce à des outils sûrs, explicables et maîtrisés.



66 Robotique et interface

- 66 Estimation de la configuration articulaire d'un préhenseur pluridigital à partir de nuages de points 3D et d'IA
- 68 Une interface de programmation cognitive
- 70 Lubrification active, frein haptique kinesthésique et vibrotactile

72 Capteurs

- 72 Réseaux de Bragg sur fibre optique pour la mesure ultrasonore à haute température
- 74 Des détecteurs quantiques lèvent le voile sur les neutrinos
- 76 Techniques de mesures électrochimiques pour la recherche de nitrites

78 Modélisation et caractérisation

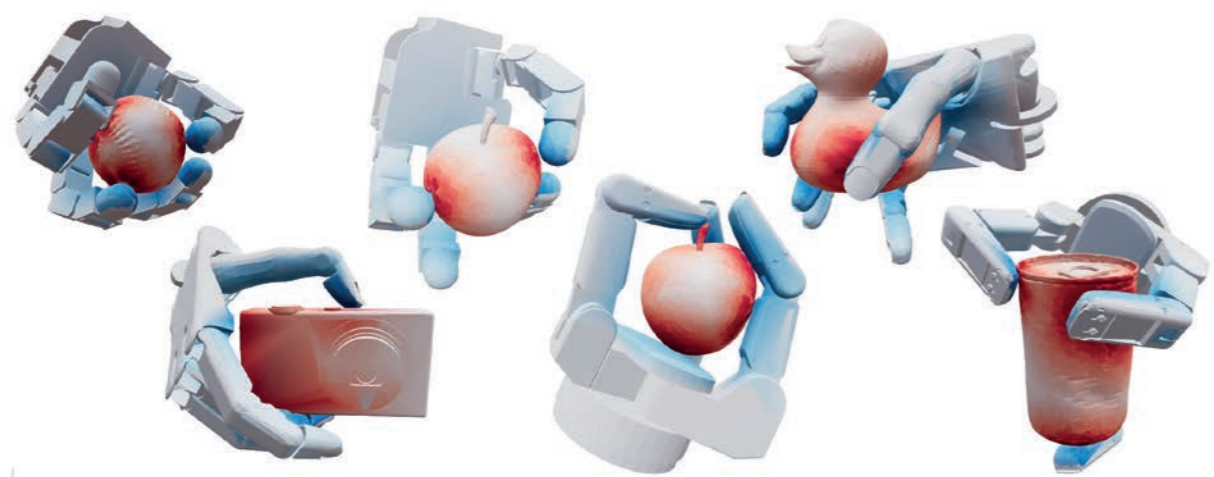
- 78 Surveillance continue par courants de Foucault de la dégradation par fatigue dans un échantillon en acier non entaillé
- 80 Algorithmes d'IA pour la spectrométrie gamma dédiée à la mesure de terrain
- 82 Photoneutrons et photofission, armes scientifiques contre les trafics

L'instrumentation numérique et l'usine du futur

Robotique et interface

Estimation de la configuration articulaire d'un préhenseur pluridigital à partir de nuages de points 3D et d'IA

La préhension robotique exige un contrôle précis du positionnement des effecteurs (bout des doigts ou phalanges) sur l'objet saisi. Pour cela, il est essentiel de connaître la configuration interne des articulations. Nous avons étudié le cas pratique où les seules informations disponibles sont le nuage de points 3D du préhenseur, généré par capteurs visuels, simulations ou réseaux neuronaux génératifs.



Exemples de saisies robotiques avec 3 préhenseurs pluridigitaux : Shadow (5 doigts), Allegro (4 doigts) et Barrett (3 doigts).



JULIEN MÉRAND
Doctorant



BORIS MEDEN
Ingénieur-chercheur



MATHIEU GROSSARD
Directeur de Recherche et expert senior

Les techniques conventionnelles de cinématique inverse (IK) ont la capacité de fournir des solutions mathématiquement exactes (si elles existent) pour déterminer la configuration des articulations à partir de la seule position de l'effecteur final (bout des doigts). Mais elles requièrent souvent une prise de décision a posteriori en tenant compte des positions de toutes les phalanges intermédiaires des doigts du préhenseur. Pour les cinématiques plus complexes, ces méthodes s'appuient sur des algorithmes d'approximation numérique qui peuvent voir leurs performances diminuer dans des environnements dynamiques.

Notre méthode propose une approche novatrice par apprentissage automatique (machine learning) grâce à un auto-encodeur variationnel conditionnel (CVAE) qui reconstruit les configurations articulaires à partir du nuage de points du préhenseur robotique.

Sur le jeu de données MultiDex, cette méthode maintient une erreur articulaire moyenne inférieure à 4% tout en assurant une inférence ultra-rapide (< 0,05 ms). Cette célérité permet de remplacer les coûteuses étapes d'optimisation numérique habituelles, rendant l'algorithme parfaitement compatible avec les contraintes du temps réel. En générant les données d'apprentissage uniquement à partir du modèle URDF, l'approche s'adapte facilement à tout type de préhenseur.

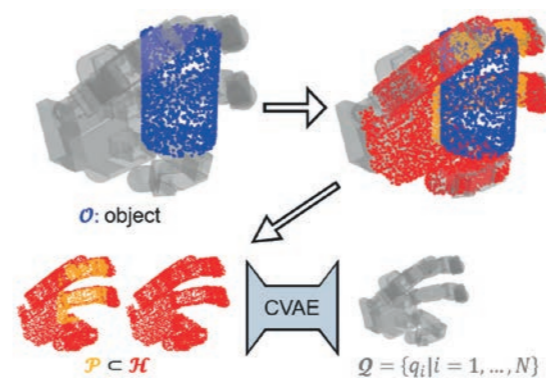
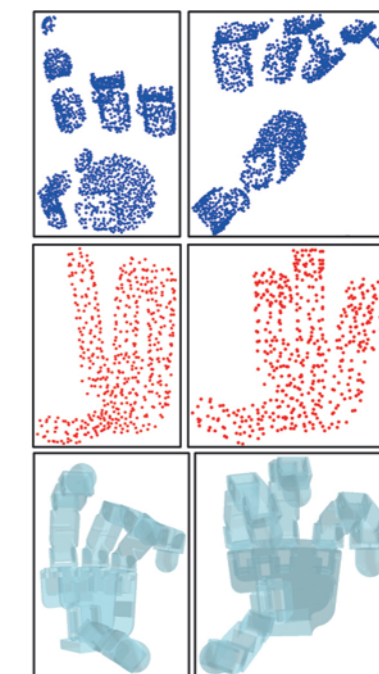


Figure 1. A partir d'un nuage de point 3D complet ou partiel d'un préhenseur robotique pluridigital, cette méthode retrouve la configuration articulaire correspondante.

4%

4% de moyenne d'erreur articulaire avec une inférence plus rapide que l'état de l'art (<0.05ms) avec le jeu de données de préhension MultiDex et la main Allegro.

Figure 2. Données utilisées : nuages de points complets ou partiels.



Publication majeure

« Leveraging CVAE for Joint Configuration Estimation of Multifingered Grippers from Point Cloud Data », J. Mérand, B. Meden and M. Grossard, 2025 IEEE 21st International Conference on Automation Science and Engineering (CASE), Los Angeles, CA, USA, 2025, pp. 895-900, <https://doi.org/10.48550/arXiv.2511.17276>

L'instrumentation numérique et l'usine du futur

Robotique et interface

Une interface de programmation cognitive

Programmer des robots est une tâche chronophage qui peut freiner leurs usages en contexte industriel, en particulier pour les PME aux tâches très changeantes. Certes, il existe des approches de programmation au niveau des tâches, mais elles manquent de contextualisation et sont réservées à des experts.

Nous avons donc développé une interface cognitive basée sur des ontologies pour offrir un outil de programmation contextuelle sans compétences robotiques préalables.

Le framework Cognitive Programming Interface (CPI) aide un utilisateur non expert à programmer un robot, en lui donnant la possibilité d'enchaîner des séquences cohérentes avec l'état courant de la scène. Cette interface repose sur trois éléments :

- 1. Une ontologie** décrivant une scène, les objets la composant (nommés acteurs, e.g. table, bouteille, pince...), leurs interfaces avec les autres acteurs (fond, zone de saisie, ou filetage pour la bouteille...) et un ensemble de skills ou compétences connues (attraper, poser, visser...);
- 2. Un interpréteur contextuel** capable de mettre à jour en temps réel l'état du monde et d'en déduire à partir de changements d'états (fermeture, ouverture de pince...) les interactions possibles entre les acteurs en fonction des interfaces que chaque acteur va cacher ou laisser disponible;
- 3. Une interface graphique** montrant la scène en 3D et les compétences réalisables à chaque étape.

Le workflow de la CPI est cyclique : le modèle du monde initialise la scène, calcule les compétences ou aptitudes possibles via un raisonneur sémantique, transmet l'ensemble des possibilités à l'interface graphique, puis se met à jour en fonction des actions choisies par l'utilisateur. Certaines étapes peuvent demander une

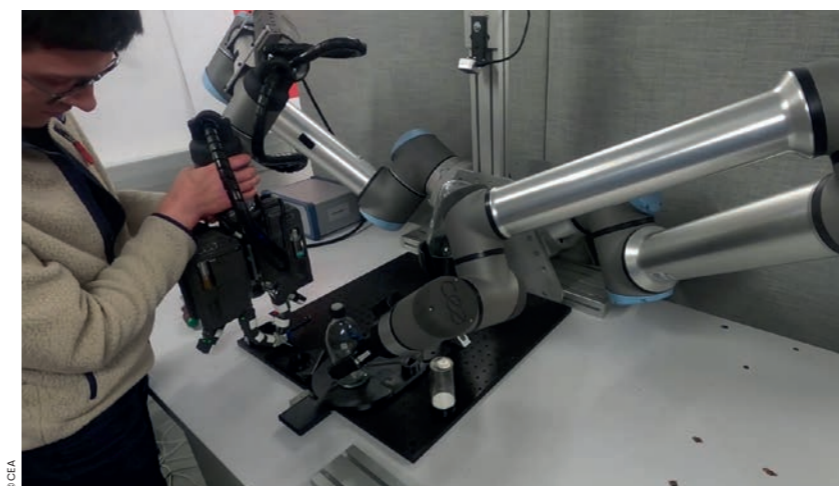
intervention humaine, notamment lorsqu'une relation ou une action particulière doit être précisée.

Dans le modèle ontologique, les acteurs sont décrits par une primitive géométrique, leurs propriétés et plus particulièrement leurs interfaces dont on connaît la position relative dans le référentiel de l'objet. Ces interfaces traduisent des capacités d'interaction (attrapable, insérable, plaçable, etc.) et permettent de raisonner non plus sur les objets mais sur leurs interactions les plus probables en fonction du contexte. Les relations entre les différentes instances de l'ontologie sont représentées sous la forme de triplets < sujet, prédicat, objet > (par exemple < robot, grasps, object >). Les capacités ou aptitudes d'interactions sont définies par un ensemble de paramètres, des préconditions ainsi que leurs effets qui permettront d'ajouter, retirer ou mettre à jour des relations dans l'ontologie. Il existe aussi des requêtes hybrides, capables de solliciter soit la base de connaissances soit l'utilisateur en cas d'ambiguïté ou d'information manquante.

Un module de reconnaissance d'interactions par démonstration est utilisé afin de déduire automatiquement, lors d'une démonstration kinesthésique (réalisée à la main par l'opérateur), la suite de compétences ou aptitudes exécutées. On fait pour cela l'hypothèse que



RAPHAËL GERIN
Ingénieur-chercheur



Apprentissage d'un geste médical au système robotique : l'opérateur joue manuellement la séquence d'opérations à automatiser (saisie du flacon par la pince à droite, prise de l'aiguille à insérer dans le capuchon du flacon avec le préhenseur dextre Abilis à gauche).

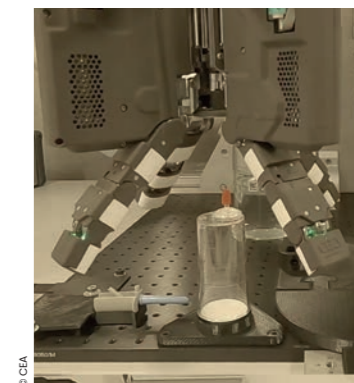


Figure 2 : Action de saisie.

chaque compétence débute ou s'achève par un changement d'état du préhenseur. Deux fonctions principales — TriggerGrasp et TriggerRelease — interprètent les mouvements, les contacts et les relations entre objets afin d'actualiser l'état sémantique du monde. Un algorithme identifie ensuite le type de compétence grâce aux prédicats modifiés, puis en déduit les paramètres exacts en comparant les changements observés avec les effets attendus des compétences possibles.

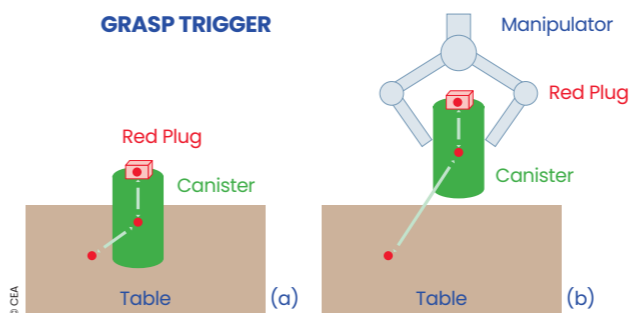


Figure 1 : Etat du système avant et après la saisie.

Publication majeure
« Cognitive Programming Interface : from Task Level Programming to Coherent Task Level Programming »
Raphaël Gerin, Julie Dumora, Olivier David, Baptiste Gradousoff, 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE), Bari, Italy, pp. 1345-1352. <https://doi.org/10.1109/CASE59546.2024.10711577>

Projet européen Tracebot
Ce projet traite de la traçabilité et de l'automatisation des processus industriels de contrôle de la stérilité dans l'industrie médicale et pharmaceutique. Deux robots UR10e ont manipulé deux conteneurs disposés sur leur support. L'expérience valide la reconnaissance automatique de compétences à partir d'une démonstration kinesthésique. Dans la démonstration, l'opérateur exécute quatre actions : la saisie puis l'insertion successive des deux conteneurs dans l'un des emplacements du support. À chaque changement d'état, le système a su correctement identifier chaque type de compétence et ses paramètres associés (objet, robot, emplacement) en comparant les changements sémantiques observés à ceux attendus. La séquence complète est reconstruite fidèlement et visualisée dans l'interface, montrant la robustesse de l'approche. Il devient alors possible de répéter cette action de manière automatique avec le robot.

L'instrumentation numérique et l'usine du futur

Robotique et interface

Lubrification active, frein haptique kinesthésique et vibrotactile

La lubrification active est un phénomène étonnant. Deux pièces solidement ancrées, par les forces de frottement, se relâchent dès que l'une d'elles est soumise à une vibration haute fréquence. Mieux encore, la force de frottement résiduelle dépend directement de la vitesse de vibration et la dynamique du système permet de commuter l'actionneur rapidement. De quoi réaliser des freins innovants particulièrement adaptés aux interfaces haptiques.

Le choix de l'actionneur est l'un des principaux verrous technologiques pour l'adoption à grande échelle des dispositifs haptiques, en particulier pour les systèmes portés tels que les gants haptiques appliqués à la réalité virtuelle. La recherche aujourd'hui est très active et s'emploie à trouver une alternative au traditionnel moteur électrique, trop encombrant. Dispositifs pneumatiques, électrostatiques ou magnétorhéologiques ont fait leur apparition dans la littérature. Ils ont comme particularité d'être des actionneurs passifs (c'est-à-dire des freins). Ils se distinguent par une grande densité d'énergie permettant de faire des systèmes puissants et compacts. La contrepartie est qu'ils ne peuvent que s'opposer au mouvement de l'utilisateur, mais cela couvre déjà une grande partie des interactions utiles en pratique. Dans ce contexte, le CEA-List étudie depuis une dizaine d'années le principe de la lubrification active et a conçu un actionneur intégré complet ouvrant la voie à un nouveau type de frein applicable à un large éventail d'applications.

Le principe est simple (voir figure 1) : une pièce mobile, typiquement une plaque en verre, est prise en étau entre deux actionneurs piézo-électriques. Lorsqu'ils sont excités à leur fréquence de résonance par un signal électrique, la pièce se libère. Il est alors possible de contrôler la force ressentie par un utilisateur déplaçant cette pièce en

modulant l'amplitude du signal d'excitation. Cette modulation peut être très rapide, jusqu'à 200Hz voire plus. Ainsi, cet actionneur est capable de rendre un effort depuis le domaine kinesthésique jusqu'au domaine tactile. Il est du reste possible de découpler les deux domaines et de contrôler indépendamment la force et la texture ressenties.

La conception et l'étude de ce frein a été le sujet des travaux de thèse de Marion Pontreau, dans le cadre du projet ANR WAVY, à l'issue desquels elle a obtenu le résultat remarquable d'une réduction de la force de

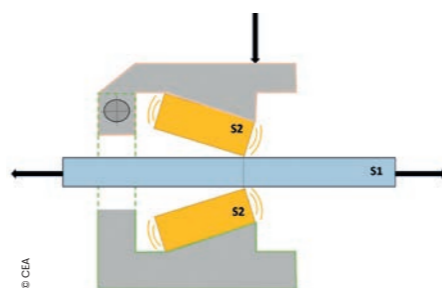


Figure 1. Principe du frein : en bleu, la pièce manipulée par l'utilisateur. En jaune, les actionneurs piézo-électriques contrôlant la résistance au déplacement. En gris, les éléments assurant la compression des piézos sur la pièce.



SYLVAIN BOUCHIGNY, Ingénieur-chercheur et expert



MARION PONTREAU, Doctorante



SABRINA PANÉELS, Ingénieure-chercheuse et experte

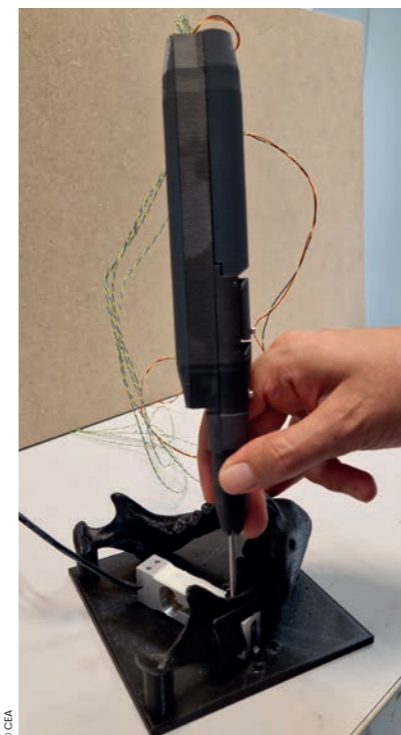
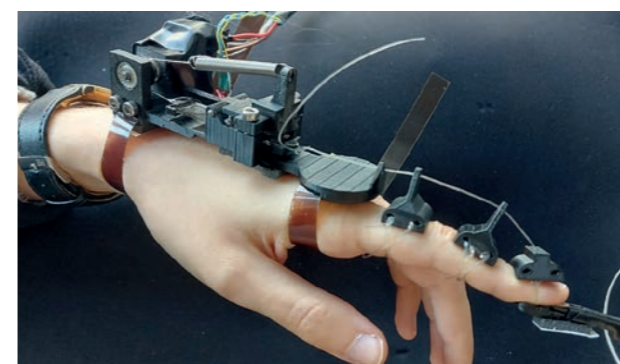


Figure 2 : Prototype de fraiseuse pour la simulation d'une ostéotomie.

frottement de plus de 90%, démontrant ainsi la pertinence de cette approche pour les applications haptiques. L'actionneur développé a été appliqué à différentes interfaces : intégré au gant haptique WAVY (voir illustration), il a permis de rendre les efforts et les textures de serrage (« Best Demo Award » à la conférence IEEE VR 2025). Il a également été utilisé pour la conception d'une fraiseuse chirurgicale destinée à la formation au geste en simulation d'une section de segment osseux (figure 2), travaux qui ont obtenu le prix « Best Student Paper » à la conférence World Haptics 2025.



Le gant haptique WAVY.

- ● Applications et cas d'usage
Cette technologie se décline aujourd'hui sur plusieurs types d'interfaces : bouton rotatif, bouton-poussoir, slider, outils ou contrôleurs portés. Il s'agit de l'une des solutions technologiques les moins chères du marché.
- ● Projets
Les travaux ont été menés dans le cadre des projets ANR WAVY et Poc In Labs SATT Paris-Saclay OTACT.
- ● Brevets
Deux déposés et un troisième en cours de rédaction :
 - FR3144344B1 S. Bouchigny, P.H Orefice, M. Pontreau, S. Paneels, Device for controlling the movement of a part
 - FR3054073A1 E. Leroy, S. Bouchigny, M. Hafez, Tangible object with improved haptic return and simplified structure and a system comprising at least one such tangible object.

- ● Publications majeures
 - « WAL: Wearable based on Active Lubrication for Virtual Reality », M. Pontreau, S. Bouchigny, S. Panéels and S. Haliyo, in *Eurohaptics 2024, LNCS, vol. 14769, pp 291-304* https://dx.doi.org/10.1007/978-3-031-70061-3_24
 - « Design and evaluation of an active lubrication brake for a surgical drilling simulator », M. Pontreau, S. Panéels, S. Bouchigny and S. Haliyo, in *World Haptic Conference, Suwon, South Korea, Jul. 2025 Best Student Paper Award* <https://doi.org/10.1109/WHC64065.2025.11123264>

90%

90% c'est la réduction typique des forces de frottement obtenue à l'aide de la lubrification active.

L'instrumentation numérique et l'usine du futur

Capteurs

Réseaux de Bragg sur fibre optique pour la mesure ultrasonore à haute température

Les environnements extrêmes, tels que les hautes températures, sont un verrou technologique important pour l'utilisation de systèmes de monitoring. Ces travaux présentent des premières mesures d'ondes ultrasonores par réseau de Bragg réalisées à plus de 1 000 °C.

Le contrôle santé intégré (*Structural Health Monitoring – SHM*) consiste en l'acquisition de données pendant l'utilisation de la structure, afin de déterminer son état de santé. Ces méthodes promettent un suivi précis de l'état de santé permettant d'optimiser la maintenance des structures. Cependant, elles présentent aussi des contraintes : l'instrumentation doit pouvoir supporter les conditions d'utilisation de la structure et ne pas perturber son fonctionnement. C'est pourquoi le CEA-List travaille sur l'élaboration de systèmes à base d'ondes guidées ultrasonores, qui se propagent sur de longues distances et sont très sensibles aux perturbations de la structure, mesurées grâce à des réseaux de Bragg sur fibre optique, peu intrusifs de par le petit diamètre de la fibre et capables de supporter des environnements extrêmes (températures extrêmes, environnement ionisant, champ électromagnétique intense...).

Ces capteurs sont sensibles notamment aux déformations et à la température. En particulier, les ondes ultrasonores sont caractérisées par de petites déformations à haute fréquence par rapport à des mesures standards avec ces capteurs, ce qui implique une technique d'interrogation spécifique nommée *edge filtering*. L'équipe a donc développé un interrogateur capable de mesurer les ultrasons malgré des variations des conditions environnementales et opérationnelles, caractérisées par une plus grande amplitude et une plus faible fréquence. Ce système a notamment été utilisé lors d'une campagne de mesures au laboratoire souterrain de Bure.

Pour tester la capacité du système à mesurer dans des environnements extrêmes, une expérience a été réalisée avec un capteur couplé à une tige métallique à l'intérieur d'un four. La tige a été chauffée à plus de 1 000°C. Une source ultrasonore, incapable de supporter de telles températures, était couplée à l'extrémité de la tige sortant du four.

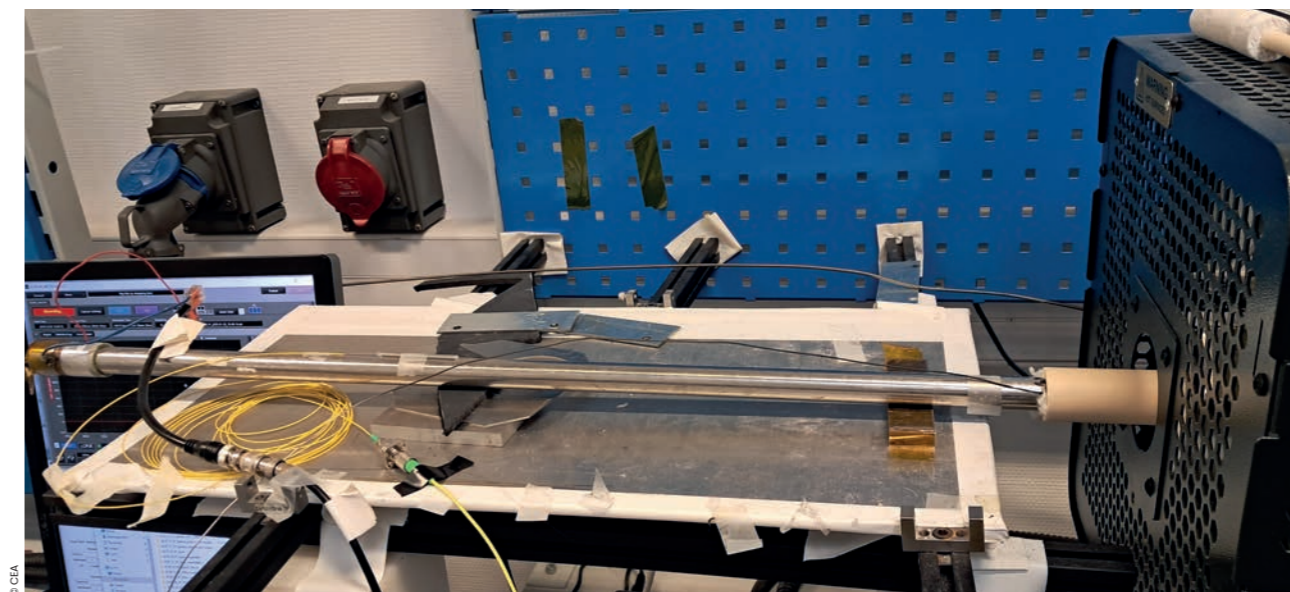
Des signaux ultrasonores ont ainsi pu être mesurés à ces températures. C'est la première fois qu'un capteur est capable de mesurer des ultrasons à de telles températures, montrant la faisabilité du monitoring par ondes guidées dans ces conditions extrêmes.



ANTOINE GALLET
Ingénieur-chercheur



ARNAUD RECOQUILLAY
Ingénieur-chercheur et expert



Expérience de mesure à haute température par réseau de Bragg. La tige métallique et la fibre optique entrent dans le four à droite, et un transducteur piézoélectrique (à gauche) permet d'émettre des ultrasons dans la tige.

Les travaux continuent afin d'optimiser la mesure, notamment la capacité du système d'interroger un plus grand nombre de capteurs simultanément sur une même fibre : actuellement, cette capacité est limitée à sept réseaux.



Figure 1 : Système de mesure Optogéro déployé à Bure.



Figure 2 : Fibre optique pour intégration dans le béton.

« Les capteurs sur fibre optique permettent un monitoring non intrusif dans des conditions extrêmes. »

••• ARNAUD RECOQUILLAY

••• Application
Monitoring à haute température (fours, aéronautique, nucléaire...)

••• Projet
Euratom FIND

••• Publications majeures
« Ultrasonic Guided Waves Measurements using Bragg Gratings in Optical Fibers under Varying Environmental Conditions ». A. RECOQUILLAY, N. ROUSSEL, L. MAURIN, et al. In: 50th Annual Review of Progress in Quantitative Non Destructive Evaluation. American Society of Mechanical Engineers, 2023. p. V001T09A009. <https://doi.org/10.1115/QNDE2023-118544>



« Ultrasonic waves detection above 1000 degree Celsius with Fiber Bragg Grating sensors ». A. GALLET, A. RECOQUILLAY, N. ROUSSEL, et al. In: 2025 IEEE Sensors Applications Symposium (SAS). IEEE, 2025. p. 1-5. <https://doi.org/10.1109/SAS65169.2025.11105178>

Mesure d'ultrasons à plus de 1 000°C

1 000°C

L'instrumentation numérique et l'usine du futur

Capteurs

Des détecteurs quantiques lèvent le voile sur les neutrinos

Particules quasi indétectables, les neutrinos ont été étudiés par la collaboration BeEST avec des capteurs quantiques ultra-sensibles, mesurant pour la première fois la taille de leur paquet d'onde à leur création. En analysant le recul du lithium-7 issu de la désintégration du béryllium-7, l'expérience révèle que le neutrino s'étend physiquement sur une zone bien plus vaste que le noyau, remettant en cause notre compréhension des désintégrations nucléaires faibles.



XAVIER MOUGEOT
Directeur de Recherche

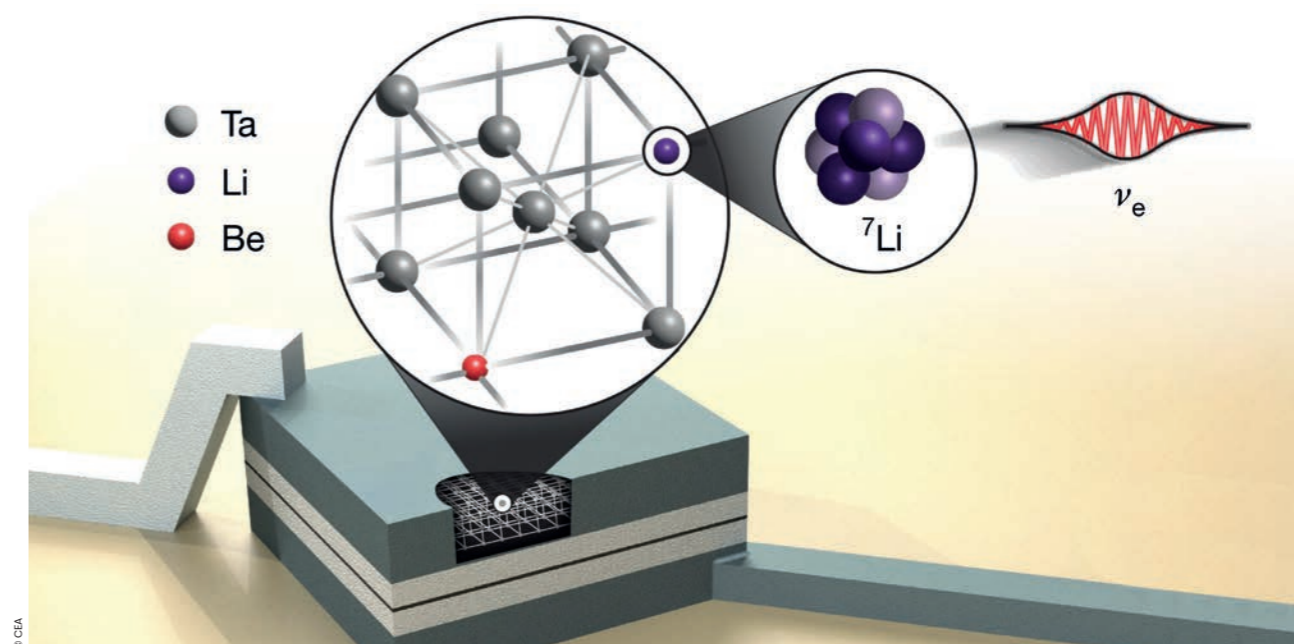
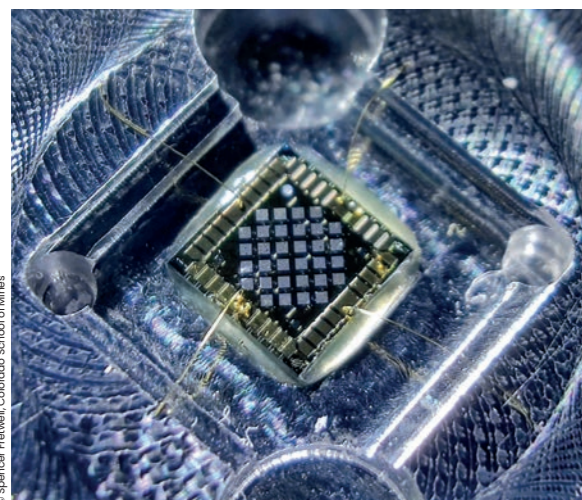


Figure extraite de Nature 638, 640-644 (2025).



Ensemble de 32 capteurs quantiques sur puce, irradiés pour implanter du béryllium-7.

Les neutrinos sont des particules fantomatiques inventées en 1930 par Wolfgang Pauli dans une tentative désespérée pour sauver le principe de conservation de l'énergie dans la désintégration bêta. Ces particules existent bel et bien mais elles interagissent extrêmement peu avec la matière. Leur étude nécessite des systèmes de détection très volumineux, comme le kilomètre-cube de glace utilisé dans l'expérience IceCube en Antarctique, et leurs propriétés demeurent mal connues. Le modèle standard de la physique des particules ne leur prédit pas de masse, mais les expériences d'oscillations ont démontré qu'ils en possèdent une. Trois types de neutrinos sont connus, et l'existence d'un quatrième, le neutrino stérile, couplé uniquement à la gravité, est encore hypothétique.

La collaboration BeEST, portée par Kyle G. Leach et dans laquelle le LNHB est impliqué depuis sa création en 2020, a développé une approche innovante : mesurer indirectement les propriétés des neutrinos avec des détecteurs extrêmement petits (environ le diamètre d'un cheveu), dans une expérience dite « sur table ». Le noyau radioactif de béryllium-7, qui se désintègre par capture d'un électron atomique, est implanté dans une jonction supraconductrice à effet tunnel, un détecteur quantique doté d'une résolution en énergie exceptionnelle. La mesure ultraprécise de l'énergie de recul du noyau, un spectre d'une centaine d'électronvolts seulement, fournit des informations sur le neutrino émis, les deux particules étant intriqués quantiquement.

« L'expérience BeEST ouvre une nouvelle fenêtre sur notre compréhension de la désintégration radioactive. »

XAVIER MOUGEOT

Comme toute particule quantique, le neutrino n'est pas strictement localisé mais possède une probabilité de présence dans une certaine région de l'espace. Jusqu'ici, il était admis qu'elle correspondait à la taille du noyau. La grande surprise révélée par l'expérience BeEST est que, dès sa création, le neutrino s'étend sur une région des milliers de fois plus grande que le noyau, bouleversant notre compréhension du processus de désintégration radioactive.

La précision de cette méthode est telle qu'elle permet également d'observer l'effet de l'environnement chimique sur le processus nucléaire. L'un des premiers résultats de la collaboration BeEST a montré que la probabilité de capture électronique du béryllium-7 peut varier d'un facteur deux entre une mesure dans un détecteur en tantale et une prédiction théorique pour un noyau dans le vide. Cette dernière a été réalisée avec le code BetaShape, développé par le LNHB depuis une quinzaine d'années et référence mondiale pour l'évaluation des données de désintégration bêta. Cet effet d'environnement chimique est actuellement le principal verrou scientifique limitant la précision des informations physiques que l'on peut extraire de l'expérience BeEST. Les travaux théoriques, complexes, se poursuivent en collaboration avec Paul-Antoine Hervieux, professeur à l'Université de Strasbourg, afin de le modéliser avec la meilleure précision possible.

Projet
La sensibilité exceptionnelle de ces capteurs quantiques permet d'envisager l'étude de l'environnement chimique sur les processus nucléaires. Ce sujet est au cœur d'un projet ERC Synergy Grant en cours d'élaboration, visant au développement d'une modélisation multi-échelle conservant l'hybridation quantique, du matériau (10^{-3} m) au noyau (10^{-15} m). Cela permettra de soustraire du spectre du béryllium-7 la composante provenant de la physique du modèle standard, ouvrant la voie à l'exploration d'une nouvelle physique, celle des neutrinos massifs, possibles candidats pour la matière noire. Cela pourrait également ouvrir la voie à la prédiction des meilleurs matériaux à utiliser pour influencer de manière contrôlée le processus de désintégration radioactif.

Publication majeure
Ce résultat majeur de la collaboration BeEST a été publié dans la revue Nature : « Direct experimental constraints on the spatial extension of a neutrino wavepacket », J. Smolsky et al., Nature 638, 640-644 (2025), <https://www.nature.com/articles/s41586-024-08479-6>



L'instrumentation numérique et l'usine du futur

Capteurs

Techniques de mesures électrochimiques pour la recherche de nitrites

Le CEA-List innove grâce au diamant, un matériau exceptionnel dont les propriétés électrochimiques remarquables sont mises à profit pour permettre de détecter la présence de nitrites dans les urines dans le contexte du diagnostic précoce de l'infection urinaire.

La détection des nitrites est une procédure de routine, utilisée notamment dans les laboratoires de biologie et les hôpitaux, pour faciliter le diagnostic des infections des voies urinaires (IVU). Un test positif aux nitrites suggère que la cause de l'infection urinaire est un organisme Gram négatif, le plus souvent *Escherichia coli*, bactérie responsable de 75 % à 95 % des infections urinaires. En effet, la présence de nitrites dans les urines est due à la conversion bactérienne des nitrates endogènes en nitrites. Elle peut être le signe d'une infection par ces types de bactéries et, par conséquent, un indicateur indirect couramment employé pour détecter une bactériurie. Actuellement, la méthode la plus courante pour détecter rapidement les nitrites utilise des bandelettes colorimétriques reposant sur la réaction de Griess, où les nitrites réagissent avec le réactif de Griess. Bien que cette méthode soit peu coûteuse, qu'elle nécessite peu de temps et d'expertise, l'interprétation du résultat peut être subjective, sa sensibilité est relativement limitée (environ 10 μM en pratique) et elle ne donne pas d'indication sur la concentration réelle des nitrites présents.

C'est pourquoi le CEA-List a travaillé en collaboration avec USENSE sur le développement d'une méthode alternative du dosage des nitrites dans les urines. USENSE est une start-up française qui développe un dispositif médical connecté et portable permettant la mesure précise et instantanée de plusieurs biomarqueurs dans l'urine.



Figure 1 : électrodes en diamant dopé au bore fabriquées au CEA-List par dépôt chimique en phase vapeur assisté par plasma microonde.



EMMANUEL SCORSONE
Ingénieur-chercheur
et expert senior



GUILLAUME LEMETAIS
CEO Usense

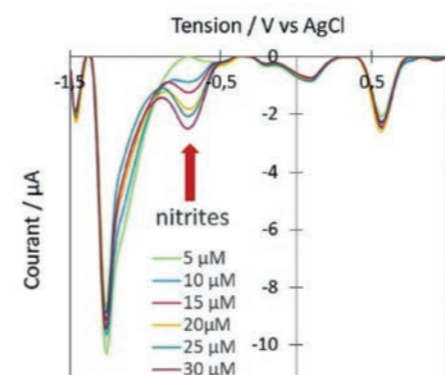


Figure 2 : voltampérométrie illustrant un exemple de réponse d'une électrode en diamant dopé au bore en présence de nitrites dans un échantillon d'urine marqué, enregistré lors de la phase de développement du procédé.

Une méthode analytique reposant sur la détection électrochimique des nitrites dans l'urine à l'aide d'électrodes en diamant dopé au bore a alors vu le jour.

Cette méthode tire avantage des propriétés remarquables des électrodes en diamant dopé au bore, fabriquées au CEA-List, à savoir une large fenêtre de potentiel en milieu aqueux, un faible courant capacitif de double couche, et une capacité d'autonettoyage de la surface des électrodes grâce à un procédé électrochimique breveté par le laboratoire il y a quelques années. Cette capacité d'autonettoyage permet en particulier la répétition des mesures un grand nombre de fois, estimé à plusieurs dizaines de milliers sans remplacer les électrodes.

Concrètement, la méthode développée repose sur une réduction électrochimique des nitrosamines, formés en particulier dans les urines à l'issue de la réaction spontanée des nitrites avec les tryptophanes provenant de la dégradation naturelle des protéines dans le corps humain.

Le protocole a été validé sur des centaines d'échantillons d'urine de patients en milieu hospitalier et laboratoire d'analyse en partenariat avec USENSE. Il permet la détection sélective des nitrites dans les urines avec un seuil de détection autour de 0,5 micromolaire.

« Le diamant, avec ses caractéristiques uniques et sa réutilisabilité quasi-infinie, possède le potentiel pour transformer le diagnostic médical. »

••• GUILLAUME LEMETAIS



••• Cas d'usage, application, transfert

Le transfert de la technologie de détection de présence des nitrites dans les urines à l'aide d'électrodes en diamant est en cours de discussion entre le CEA-List et USENSE. Ces derniers visent une commercialisation de leur dispositif « Jimini » d'analyse d'urine intégrant cette technologie dans un horizon proche. De manière générale, cet outil vise à remplacer les tests bandelettes pour permettre une alternative de dosage plus fiable, plus sensible et numérisable des nitrites dans l'urine. Un cas concret d'application repose sur l'arrêté du 28 novembre 2023 qui autorise les pharmaciens à réaliser des tests urinaires donnant lieu à une prise en charge par l'assurance maladie. Ce test comprend la recherche, a minima, de nitriturie et de leucocyturie, dans le cadre d'un bilan de symptôme(s) évocateur(s) d'une cystite aiguë non compliquée chez la femme.

••• Brevet

Le procédé de détection des nitrites dans les urines à l'aide d'électrode diamant a fait l'objet d'un dépôt de brevet en 2024

••• Partenaire majeur

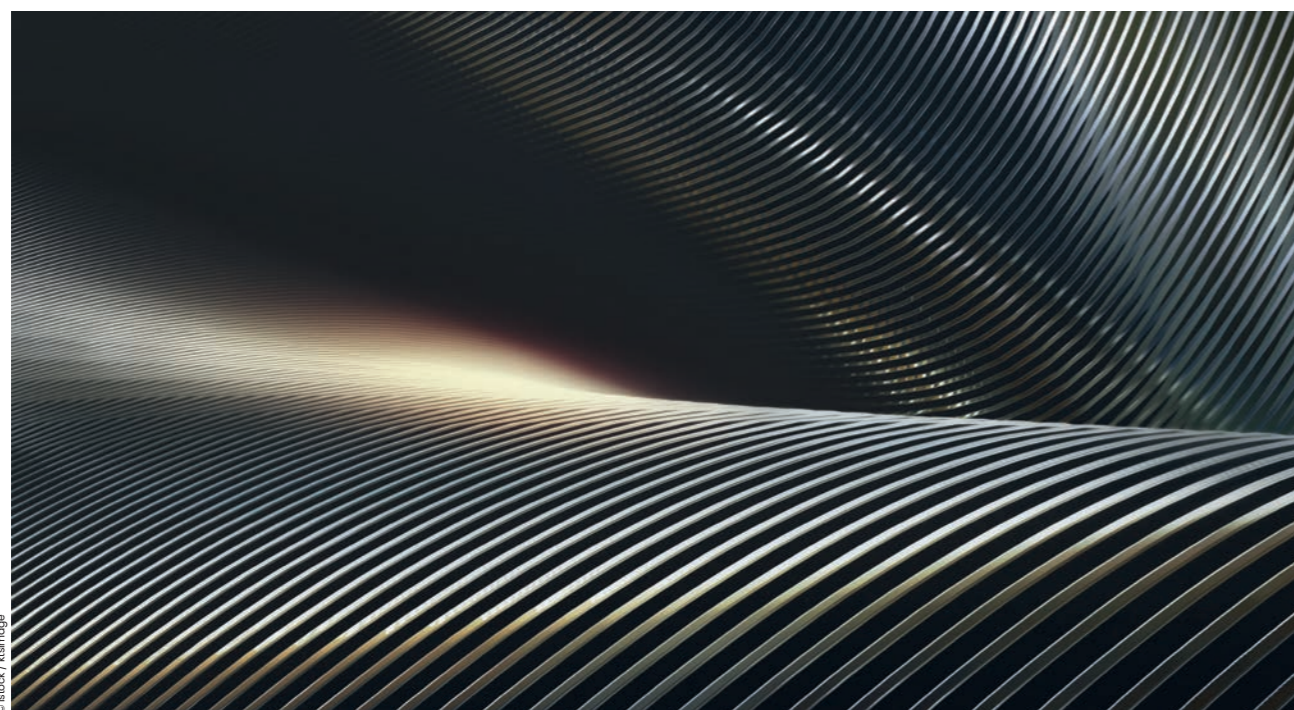
Société USENSE



L'instrumentation numérique et l'usine du futur

Modélisation et caractérisation

Surveillance continue par courants de Foucault de la dégradation par fatigue dans un échantillon en acier non entaillé



La fatigue des matériaux métalliques est un phénomène critique dans de nombreuses applications industrielles, entraînant une dégradation progressive des propriétés mécaniques. La surveillance de cette dégradation est essentielle pour assurer la sécurité et la durabilité des structures. Les méthodes électromagnétiques sont particulièrement adaptées pour cette surveillance en raison de leur sensibilité aux changements microstructuraux.

L'étude a été menée sur un échantillon en acier structural de grade C22E/C22R, avec une section transversale minimale de 7,6 mm et une longueur de jauge de 12 mm. L'échantillon a été soumis à un chargement cyclique de 230 MPa à une fréquence atteignant 5 Hz, jusqu'à la rupture. Les mesures par courants de Foucault ont été effectuées à différentes fréquences d'excitation (50, 100, 250 et 500 kHz) pour surveiller les variations de la perméabilité magnétique.

Un modèle semi-analytique a été utilisé pour résoudre le problème direct, c'est-à-dire calculer l'impédance en fonction de la perméabilité magnétique. Les résultats expérimentaux ont été comparés aux simulations pour valider le modèle. La perméabilité relative μ du matériau a été estimée dans l'intervalle 150-180, avec une perméabilité décroissante en fonction du temps au cours de l'expérience.

Les résultats ont montré une différence évidente entre la partie basse fréquence (en dessous de 3000 cycles de fatigue) et la partie haute fréquence des mesures. Cette différence est associée à différents mécanismes microstructuraux qui se produisent à différents moments de l'expérience. La partie haute fréquence des mesures a démontré un comportement presque linéaire, ce qui a été confirmé par l'analyse des valeurs de tension réelles.

Les mesures par courants de Foucault ont été complétées par des mesures de bruit magnétique Barkhausen (MBN) et des données thermographiques pour corrélation. Les variations de température ont été prises en compte pour corriger les dérives liées au chauffage de l'échantillon.

Les résultats obtenus indiquent que les mesures par courants de Foucault peuvent être utilisées pour surveiller la dégradation par fatigue dans les matériaux métalliques. La variation de la perméabilité magnétique peut être corrélée avec les changements microstructuraux, ce qui permet une évaluation non destructive de l'état du matériau.

Cette étude démontre le potentiel des mesures par courants de Foucault pour la surveillance de la dégradation par fatigue. Les résultats ouvrent la voie à des applications industrielles pour la surveillance en temps réel de l'état des matériaux, contribuant ainsi à la sécurité et à la durabilité des structures.

Publication majeure
 « Continuous eddy-current monitoring of fatigue degradation in an unnotched steel specimen ». Skarlatos A, Boller C, Lyamkin V, et al. *International Journal of Applied Electromagnetics and Mechanics*. 2025; <https://doi.org/10.1177/13835416251345615>

Cas d'usage, application, transfert
 La surveillance continue de la dégradation par fatigue dans les matériaux métalliques est essentielle pour assurer leur sécurité et leur durabilité. Les mesures par courants de Foucault, combinées à des modèles semi-analytiques, offrent une méthode efficace pour évaluer les changements microstructuraux et surveiller l'état des matériaux. Les résultats de cette étude montrent que cette approche peut être utilisée pour une surveillance en temps réel, ouvrant ainsi de nouvelles perspectives pour des applications industrielles.

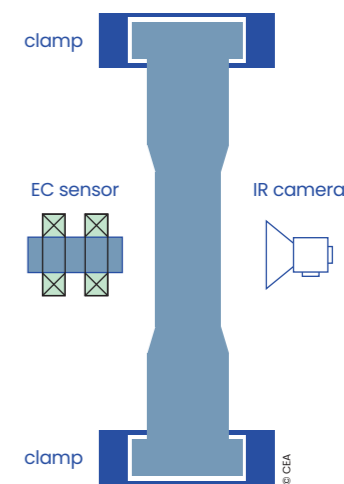


Figure 1 : montage des capteurs.

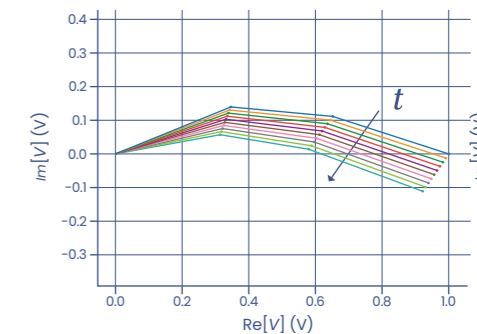


Figure 2 : évolution temporelle des signatures dans le plan complexe.

L'instrumentation numérique et l'usine du futur

Modélisation et caractérisation

Algorithmes d'IA pour la spectrométrie gamma dédiée à la mesure de terrain

En spectrométrie gamma, il existe une forte demande pour des algorithmes automatiques assurant une prise de décision rapide pour des mesures *in situ* à faible statistique. D'autres contraintes doivent également être prises en compte : une opérationnalité pour des détecteurs de faible résolution en énergie et une gestion robuste des fausses alarmes, notamment dans le cas d'un spectre déformé.

Pour répondre à ces enjeux, un algorithme de démixage spectral hybride (*machine learning* et statistique), appelé SEMSUN, a été développé en utilisant un auto-encodeur interpolatoire (IAE) permettant l'apprentissage des déformations spectrales induites par des phénomènes d'atténuation et de diffusion Compton dans l'environnement d'une source. L'auto-encodeur est un réseau de neurones ayant la propriété d'extraire les caractéristiques d'un spectre dans une représentation compressée (encodeur) ; il peut être utilisé comme modèle génératif pour reconstruire les spectres à partir des caractéristiques apprises (décodeur). La technique du démixage spectral repose sur un modèle de mélange de signatures spectrales propres à chaque radionucléide présent. Le code SEMSUN utilise le modèle IAE pour imposer une contrainte sur la déformation de ces signatures spectrales dans la résolution d'un problème inverse.

Comparée à un état de l'art basé sur l'apprentissage de bout en bout, cette approche hybride permet d'optimiser l'identification et la quantification des radionucléides en intégrant explicitement les caractéristiques physiques d'une mesure.



CHRISTOPHE BOBIN
Ingénieur-chercheur
et expert senior

Pour établir la preuve de concept, l'entraînement du modèle IAE s'appuie sur des simulations des interactions rayonnement-matière avec le code Geant4, dans une configuration géométrique où une source ponctuelle est placée au centre d'une sphère (acier, plomb), avec un détecteur NaI(Tl) de 3"×3" (voir figure 1). Les résultats de la reconstruction des spectres montrent que le modèle IAE capturent efficacement la déformation spectrale^[1].

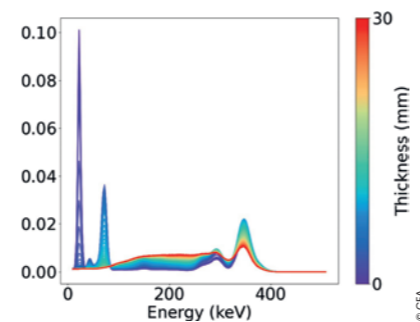


Figure 1 : Déformation du spectre d'une source de Ba-133 placée au centre d'une sphère en acier, en fonction de l'épaisseur de la sphère (simulations Geant4).

Pour l'identification automatique des radionucléides, l'algorithme MoSeVa^[2] a été développé en combinant le code SEMSUN avec une technique de sélection de modèle basée sur un test du rapport de vraisemblance. La présence d'un radionucléide est déterminée à l'aide d'un critère de parcimonie appliqué aux différentes signatures spectrales regroupées dans une librairie. Les tests ont été effectués sur différents mélanges (jusqu'à 4 radionucléides dans une librairie de 12 signatures spectrales) avec le mouvement propre. Différents comptages dans les spectres ont également été étudiés (jusqu'à un comptage minimum de 2500 événements ; voir figure 2).

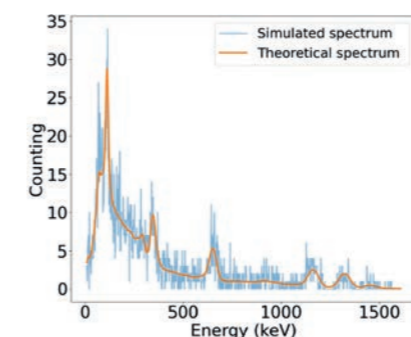


Figure 2 : Spectre de mélange de Co-57, Co-60, Ba-133 et Cs-137 avec le mouvement propre, simulé pour un total de 2500 événements.

Une comparaison a été réalisée entre le démixage spectral hybride et un état de l'art basé sur l'apprentissage de bout en bout avec des réseaux de neurones convolutifs utilisés comme classifieur ou pour des applications de régression. Les résultats confirment que le démixage spectral hybride offre de meilleures performances à faible statistique en termes d'identification et de robustesse, avec le taux de fausses alarmes attendu. Le démixage spectral hybride permet également d'obtenir une quantification plus précise au-delà des faibles statistiques^[3]. Le benchmark GammaBench (incluant des spectres, des métriques d'évaluation, etc.) offre la possibilité de comparer les performances de nouveaux développements avec le démixage spectral hybride.

**×10
SUR LA
JUSTESSE**

C'est le gain réalisé grâce au démixage spectral hybride sur la quantification face aux approches *end-to-end*.

« Le démixage spectral hybride offre une solution adaptée à la problématique de la mesure *in situ* à faible statistique sensible aux variabilités spectrales »

• • • CHRISTOPHE BOBIN

• • • Application, cas d'usage, transfert

Mesures environnementales *in situ* suite à un accident nucléaire ou radiologique avec rejet, surveillance du trafic illicite des matières radioactives ou encore caractérisation radiologique pour l'assainissement et le démantèlement des installations nucléaires, etc.

• • • Publications majeures



[1] « A hybrid Machine Learning unmixing method for automatic analysis of gamma-spectra with spectral variability ». Phan D.T., et al., *Nuclear Inst. and Methods in Physics Research A 1060 (2024) 169028*.
<https://doi.org/10.23919/EUSIPCO63174.2024.10715033>



[2] « Automatic identification and quantification of γ -emitting radionuclides with spectral variability using a hybrid Machine Learning unmixing method ». Phan D.T., et al., *Radiation Physics and Chemistry 232 (2025) 112654*.
<https://doi.org/10.1016/j.radphyschem.2025.112654>



[3] « Comparative study of machine learning and statistical methods for automatic identification and quantification in γ -ray spectrometry ». Phan D.T., et al., *Nuclear Inst. and Methods in Physics Research A 1083 (2026) 171088*.
<https://doi.org/10.48550/arXiv.2508.08306>

L'instrumentation numérique et l'usine du futur

Modélisation et caractérisation

Photoneutrons et photofission, armes scientifiques contre les trafics

Le contrôle de conteneurs maritimes constitue un enjeu crucial pour la sécurité intérieure de l'Europe. Adossé à un savoir-faire de plus de trente ans dans le domaine des méthodes de mesures nucléaires non destructives actives, le CEA-List a récemment développé deux approches en rupture visant à améliorer la détection de drogues d'une part, de matière nucléaire d'autre part.

Afin de lutter contre le trafic illicite de matières dangereuses, les équipes du CEA-List développent aujourd'hui deux innovations majeures dans le domaine des méthodes de mesures nucléaires non-destructives actives. L'une associe l'utilisation d'un accélérateur linéaire d'électrons (linac), la spectrométrie des photoneutrons et l'intelligence artificielle pour identifier des éléments légers, caractéristiques des explosifs et stupéfiants. L'autre, portée dans le cadre du projet européen MULTISCAN 3D avec le LMU^[1]-CALA^[2] de Munich, ouvre la voie à une rupture technologique dans la détection de matière nucléaire.

Dans le champ des substances illicites, la méthode mise au point s'appuie sur un linac générant un rayonnement de freinage capable d'induire des réactions photonucléaires. Les neutrons émis, dont le spectre porte la signature des éléments présents (carbone, azote, oxygène), sont détectés grâce à des scintillateurs liquides permettant de discriminer efficacement les neutrons d'intérêt au sein d'un intense flash photonique. Si ces spectres présentent des structures riches mais souvent déformées par la géométrie des objets contrôlés et la réponse des détecteurs, l'analyse automatique vient lever cet obstacle. Le modèle d'apprentissage profond DeepNSI, fondé sur des réseaux neuronaux convolutionnels entraînés par simulation Monte-Carlo, parvient en effet à identifier les éléments même à de très faibles concentrations (moins de 4 % pour l'azote). D'autres algorithmes, comme le réseau élastique non négatif, permettent en complément d'estimer leurs abondances relatives. Cette approche offre ainsi une identification chimique avancée, robuste et rapide et permet d'envisager une réponse aux attentes liées au contrôle de fret.

En parallèle, pour répondre à la menace liée au trafic de matière nucléaire, le CEA-List a réalisé une première mondiale en démontrant expérimentalement la photofission d'uranium appauvri à l'aide d'une source photonique entièrement nouvelle. Celle-ci repose sur la diffusion Compton inverse (ICS) : un faisceau laser femtoseconde accélère des électrons à très haute énergie dans un



VALENTIN BLIDEANU
Ingénieur-chercheur et directeur de recherche



ADRIEN SARI
Ingénieur-chercheur et expert senior

plasma, produisant ensuite des photons suffisamment énergétiques pour initier la photofission. Pour la première fois, une telle source se substitue à un linac traditionnel. Les caractéristiques quasi mono-énergétiques du faisceau, son potentiel d'intégration et son caractère modulable ouvrent des perspectives majeures pour le déploiement futur de mesures nucléaires actives visant à détecter des actinides au sein de conteneurs maritimes.

Ces deux avancées, l'une tournée vers les substances illicites, l'autre vers les matières nucléaires, illustrent une même dynamique : l'émergence de technologies de rupture, plus sensibles et plus intelligentes, capables d'élever significativement le niveau de sécurité des infrastructures portuaires européennes.

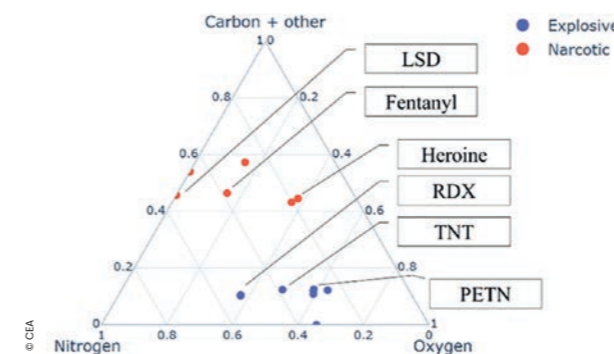


Figure 1 : Graphiques ternaires pour plusieurs molécules d'intérêt en fonction de leurs ratios de carbone, d'oxygène et d'azote construits sur la base des photoneutrons dans le spectre total.

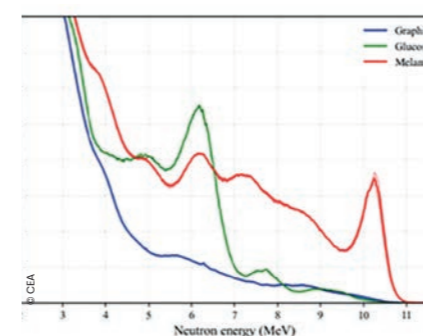


Figure 2 : Spectres des photoneutrons mesurés pour trois matériaux : graphite (riche en carbone), glucose (riche en oxygène) et mélamine (riche en azote) permettant d'identifier à travers les structures présentes les signatures spécifiques de ces éléments.

« L'IA analyse les spectres photoneutro- niques déformés et en extrait les signatures élémentaires que la seule spectrométrie ne permet pas d'isoler. »

• VALENTIN BLIDEANU

« En exploitant une source ICS fondée sur un laser, nous franchissons une étape clé dans le développement de technologies avancées de détection de matière nucléaire. »

• ADRIEN SARI

• Cas d'usage, application, transfert
DeepNSI ouvre la voie à l'identification d'éléments légers dans des spectres photoneutro- niques complexes, pour la détection de matériaux illicites, la radioprotection ou le démantèlement. Les perspectives de transfert industriel des travaux reposant sur la spectro- métrie des photoneutrons sont à l'étude, tandis que la source laser-plasma permet d'explorer de nouveaux régimes pour la photofission.

• Brevets
• IPA-SN (Dispositif et procédé de détection d'une substance particulière dans un objet par interrogation photonique active) : **FR2112182**
• IDepNSI (méthodes d'analyse avancées des spectres photoneutro- niques) : **FR2210022**

• Projet et/ou partenariat majeur
Projet européen MULTISCAN 3D ; collaboration LMU-CALA ; montée en TRL au sein de Doséo ; poursuite des travaux sur la source laser-plasma dans le cadre de la thèse de Johann Piekar.

• Publication majeure
« DeepNSI: Element identification in experimental photoneutron spectra for illicit material detection », C. Besnard-Vauterin, V. Bliedeanu, B. Rapp, *Applied Radiation and Isotopes* 225 (2025) 112014; article sur l'utilisation de la source laser-plasma en cours de rédaction. <https://doi.org/10.1016/j.apradiso.2025.112014>

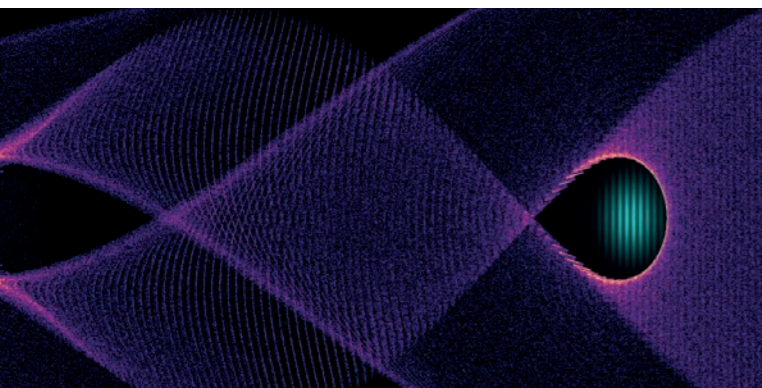


Illustration du processus d'accélération laser-plasma

[1] LMU : Ludwig-Maximilians-Universität München.
[2] CALA : Centre for Advanced Laser Applications.

Selon l'European drug report 2024, près de 70% des saisies réalisées par les autorités douanières de l'UE ont lieu dans les ports, notamment dans des conteneurs maritimes

70%

05

Cybersécurité et numérique de confiance

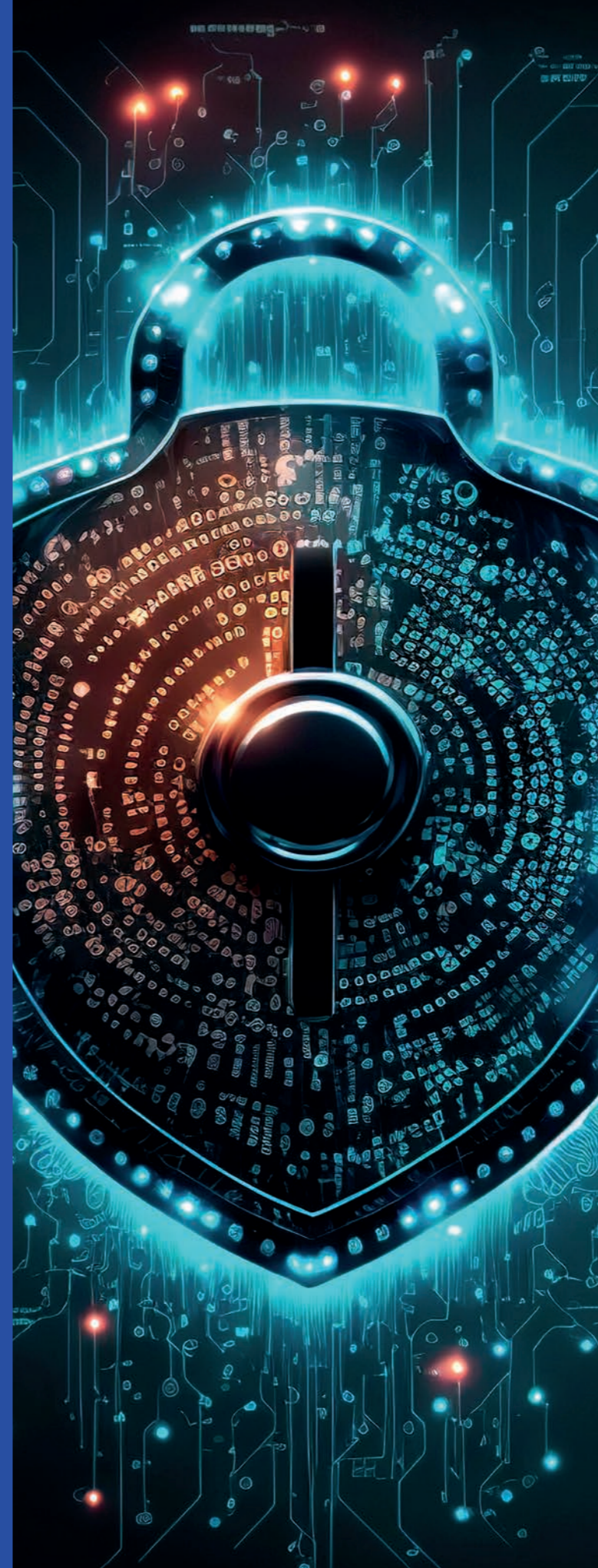
Avec la complexité et l'omniprésence logicielles et l'essor du quantique, les besoins en cybersécurité évoluent rapidement. De l'identification des vulnérabilités, leur priorisation à la préparation de défenses cryptographiques, bâtir des systèmes fiables exige de relier théorie et réalités opérationnelles.

Les logiciels modernes font face à un défi croissant en matière de complexité, ainsi qu'à la confiance dans leur capacité à résister aux attaques. L'analyse statique par méthodes formelles permet de vérifier mathématiquement l'absence de classes entières de vulnérabilités, comme les dépassements de tampon largement exploités par les attaquants. Les approches classiques imposaient une réanalyse complète après chaque

modification, incompatibles avec l'intégration continue. Des avancées récentes en analyse statique formelle réutilisent les résultats précédents, réduisant drastiquement les temps d'analyse tout en conservant des garanties mathématiques, rendant ces méthodes compatibles avec le développement agile de systèmes critiques.

Cependant, la détection automatique de vulnérabilités dépasse désormais la capacité de correction. Chaque année, des milliers de vulnérabilités sont associées dans les bases de données CVE au noyau Linux en raison de sa taille, de sa complexité et de la quantité de code qu'il contient. Mais ce n'est pas un cas isolé : face à toutes ces vulnérabilités identifiées, lesquelles traiter en priorité ? L'évaluation de l'exploitabilité, fondée sur le contrôle de l'attaquant, apporte la réponse. De nouvelles techniques analysent les valeurs exploitables des paramètres, permettant un classement automatique des failles réellement critiques, validé sur des CVE réels.

Enfin, l'informatique quantique menace le chiffrement actuel. Le risque « collecter aujourd'hui, déchiffrer demain » est réel. La cryptographie post-quantique et le chiffrement totalement homomorphe offrent des protections durables, rendues pratiques grâce à l'accélération matérielle, avec des gains de performance atteignant 94 à 96 %.



86 Détection de vulnérabilités

- 86 Vérification incrémentale de logiciels, pour la cybersécurité
- 88 Contrôle d'attaquant et priorisation de bugs

90 Technologies de sécurisation

- 90 Chiffrement rapide et homomorphe complet
- 92 Accélérateur étroitement couplé à un cœur RISC-V pour la cryptographie post-quantique

Cybersécurité et numérique de confiance

Détection de vulnérabilités

Vérification incrémentale de logiciels, pour la cybersécurité

En cybersécurité, la détection immédiate de vulnérabilités logicielles nouvellement introduites est essentielle. L'intégration récente dans Frama-C de techniques génériques d'analyse incrémentale permet de vérifier rapidement un code après des modifications localisées. Les premiers résultats sont très prometteurs en termes de temps d'analyse, tout en conservant la fiabilité.

Lorsqu'on cherche à garantir l'absence de vulnérabilités dans un code source, il ne suffit pas de disposer d'un outil capable de les détecter : on souhaite s'assurer qu'il n'en existe aucune.

C'est un des objectifs de l'outil de vérification formelle Frama-C du CEA-List : vérifier automatiquement, à l'aide de preuves mathématiques, l'absence d'une large classe de vulnérabilités, telles que les *buffer overflows* (dépassements de l'emplacement mémoire réservé pour des données, dus à l'absence de précautions lors de la programmation).

Cette vérification demande un certain investissement, mais est d'autant plus facile à mettre en œuvre qu'elle est intégrée tôt dans le cycle de développement. Cela permet de détecter l'introduction de défauts à un moment où ils restent faciles à corriger. Cela permet également de mettre en évidence, très en amont, les contraintes de l'analyse et d'inciter les développeurs à adopter un style de programmation mieux adapté à la vérification automatique.

L'utilisation de Frama-C s'intègre idéalement dans un processus d'intégration continue, pratique de développement dans laquelle le code est testé et validé à chaque modification, dès le début du projet. Pour que Frama-C puisse être intégré à un tel processus, son temps d'ana-

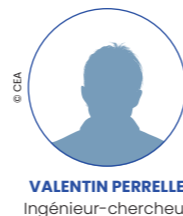
lyse doit rester comparable à l'intervalle entre deux modifications du code. Pour répondre à cette contrainte, nous avons développé dans Frama-C des techniques d'analyse incrémentale.

Le principe consiste à repartir d'une analyse déjà effectuée et à réutiliser autant que possible ses résultats afin d'accélérer l'analyse d'une version modifiée du code. La difficulté majeure est d'établir formellement que cette réutilisation ne compromet pas la correction : les résultats doivent rester aussi fiables que si l'analyse était complète.

Deux techniques ont été mises en œuvre.

La première repose sur la réutilisation de résumés de fonctions du code que l'analyseur a produit lors de la première analyse. Ces résumés correspondent à des propriétés mathématiques déduites de chaque fonction. En comparant les versions du code et en explorant leur structure, on détermine quels résumés peuvent être réutilisés pour la nouvelle analyse de la fonction.

La seconde technique vise à accélérer la recherche des invariants de boucle (propriétés logiques du code) : les invariants découverts lors d'une analyse précédente sont réutilisés comme points de départ pour trouver plus rapidement les invariants de la nouvelle version du code.



« Un usage précoce de Frama-C dans le développement logiciel permet de détecter les vulnérabilités au moment où elles sont faciles à corriger. »

VALENTIN PERRELLE

Les premiers résultats expérimentaux, à partir de codes open source dont l'historique de modifications est public, montrent que l'analyse incrémentale de code avec Frama-C permet de réduire le temps d'analyse d'un facteur de $\times 1,5$ à $\times 28$, avec une moyenne de $\times 7,83$. L'expérimentation met également en évidence une très faible variation des résultats de l'analyse, sans remettre en cause sa fiabilité. Cela confirme ainsi la pertinence de notre approche dans un contexte d'intégration continue.

Publications majeures

Ces travaux ont été réalisés en partenariat avec Sorbonne Université et s'inscrivent dans le cadre du projet SecurEval du PEPR Cybersécurité, financé par France 2030.



[1] « Reusing Caches and Invariants for Efficient and Sound Incremental Static Analysis » M. Razafintsialonina, D. Bühler, A. Miné, Valentin Perrelle, J. Signoles ; à *ECOOP 2025 (rang A)*
<https://doi.org/10.4230/DARTS.11.2.15>



[2] « Réutilisations de caches et d'invariants pour l'analyse statique incrémentale » aux *JFLA 2024*
<https://inria.hal.science/hal-04406463>

Figure 1 : Réutilisation d'une partie de l'analyse pour analyser le code modifié.

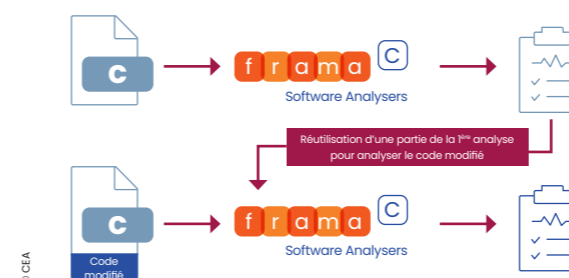
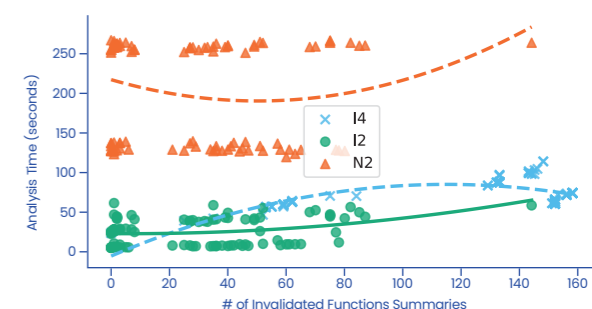


Figure 2 : Gain de temps des analyses incrémentales (vert) par rapport aux analyses complètes (orange) sur cent codes cryptographiques open-source, avec en abscisse le nombre de résumés de fonctions non réutilisables.



Cybersécurité et numérique de confiance

Détection de vulnérabilités

Contrôle d'attaquant et priorisation de bugs

La recherche automatique de bugs constitue un succès majeur de la recherche en cybersécurité des 15 dernières années, et en particulier de l'analyse automatique de code. Fuzzing, exécution symbolique et interprétation abstraite, entre autres techniques, permettent de découvrir des bugs dans nos programmes à un rythme sans précédent. Mais une question se pose alors : *que faire de tous ces bugs ?*

En observant les dizaines de milliers de bugs découverts par les plateformes publiques de fuzzing (test aléatoire) comme OSS-Fuzz (Google), il est clair que les capacités de correction de la plupart des équipes de développement ne sont plus à la hauteur. Hormis leur quantité, la clarté des rapports fournis par ces méthodes automatiques en matière d'impact sécuritaire laisse fortement à désirer, ce qui rend pratiquement impossible la *priorisation des bugs les plus dangereux* en vue de leur correction. La nécessité de développer des techniques automatiques d'évaluation de vulnérabilité et d'exploitabilité se fait donc plus que jamais sentir.

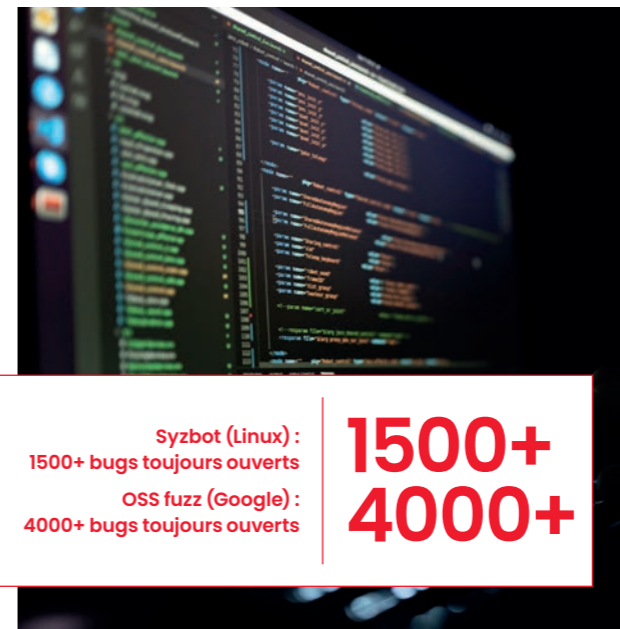
Nos travaux introduisent une formalisation du concept de contrôle de l'attaquant sur une vulnérabilité donnée : *à quel degré un attaquant peut-il influencer le comportement d'un certain type de bug au travers des entrées du programme ?* Intuitivement, un bug dont les effets peuvent être finement contrôlés par l'attaquant est plus

dangereux qu'un bug qui offrirait des possibilités moindres à l'attaquant. Il s'agit donc d'évaluer pour un même bug la *quantité* de comportements uniques contrôlables par l'attaquant. Mais au-delà de cette simple dimension quantitative, il est également nécessaire de distinguer les différents niveaux de dangerosité de ces comportements et d'intégrer toutes ces informations en un score final unique. Là réside toute la nouveauté et la complexité du problème. Le choix d'employer des méthodes formelles dans ce but garantit la transparence et l'explicabilité des résultats, tout en ouvrant de nouvelles perspectives de recherche dans un domaine jusque-là peu abordé, et où les rares travaux préexistants se fondent toujours *in fine* soit sur le jugement préalable d'un expert, soit sur des classifications grossières *a priori* (assigner la même dangerosité à toute une classe de bugs).

Ces travaux ont donné lieu à une publication à Usenix Security 2025, une des conférences scientifiques majeures en sécurité (rang A*) ainsi qu'à un nouveau plugin de la plateforme d'analyse de code binaire BINSEC du CEA-List. Notre approche a été évaluée et validée sur des tâches d'évaluation de vulnérabilités réelles de type corruption de mémoire issues de bases de CVEs (liste de vulnérabilités logicielles) ainsi que d'un benchmark de référence composé de vulnérabilités découvertes par Google OSS-Fuzz. Les programmes ciblés varient de bibliothèques largement utilisées



SÉBASTIEN BARDIN
Ingénieur-chercheur
et expert fellow



comme *libxml2* ou *libpng* à des systèmes complexes de grande échelle comme *openssl* ou le bootloader universel *u-boot*, où notre approche a démontré sa capacité à proposer une classification pertinente de la dangerosité des bugs, au-delà de l'état de l'art actuel.

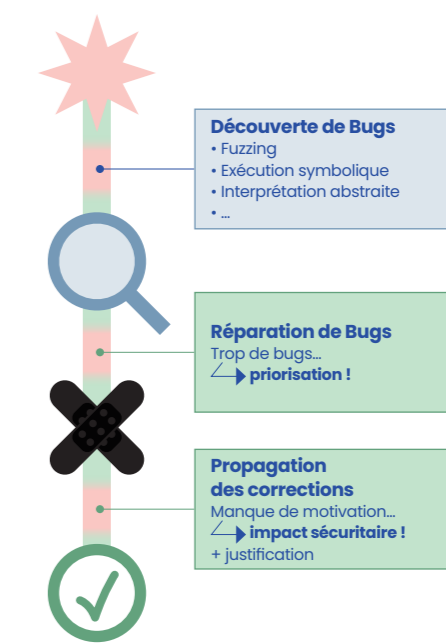


Evolution du nombre de bugs découverts par fuzzing (test aléatoire) dans Linux et toujours ouverts.

• • > **Projet et/ou partenariat majeur**
PEPR Cybersécurité/REV. Intérêt de l'ANSSI et de Schneider.

• • > **Publication majeure**
« **Attacker Control and Bug Prioritization** ». G. Lacombe et S. Bardin. *USENIX Security 2025. Flagship conference in system security (A*)*
<https://doi.org/10.48550/arXiv.2501.17740>

• • > L'évaluation du contrôle d'attaquant sur des bugs ne constitue qu'un premier pas vers l'évaluation automatique de vulnérabilités. De nombreuses problématiques restent en effet à traiter. Comment évaluer d'autres aspects de l'exploitabilité de bugs ? Comment recombinaison des résultats d'analyses disparates en un unique score et en valider la pertinence ? Enfin, comment déployer ces analyses à grande échelle avec un temps de calcul suffisant pour une évaluation à la volée ? Nos travaux marquent un premier pas fondateur, et nous explorons dans le cadre du PEPR Cybersécurité à la fois des extensions conceptuelles et une mise en application réelle.



Notre méthode de priorisation des bugs.

« Il est impossible de corriger tous les bugs repérés par les outils de vérification. Par où commencer ? Nous utilisons les méthodes formelles pour les prioriser selon leur dangerosité. »

• • • SÉBASTIEN BARDIN

Cybersécurité et numérique de confiance

Technologies de sécurisation

Chiffrement rapide et homomorphe complet

La confidentialité est un enjeu majeur pour de nombreux domaines (militaire, santé, industriel, financier). Le calcul sur données chiffrées, réalisable grâce au chiffrement homomorphe (FHE), est une piste prometteuse mais il reste encore de nombreux défis à relever, sur lesquels le CEA-List s'investit depuis des années. En 2025, nous avons proposé un processeur FHE virtuel, nom de code MadPanthera, manipulant des chiffres d'entiers de 8 bits, avec des résultats dépassant l'actuel état de l'art.

En 2020, après Cingulata, un compilateur open-source pour le FHE développé par le CEA-List, l'équipe était à la recherche d'une voie pour donner un « second souffle » de performance à l'approche FHE dite généraliste, permettant de réaliser la promesse de réaliser n'importe quel algorithme dans le domaine chiffré, mais aussi d'identifier des pistes qui se démarqueraient de l'état de l'art en vue d'une accélération HW.

Une occasion s'est présentée en 2022 autour de la thèse de Daphné Trama. La stratégie a été de définir un petit corpus d'opérateurs FHE basés sur TFHE et permettant de reconstruire « ex nihilo » une approche de FHE à vocation généraliste plus efficace que l'état de l'art. Concrètement, l'objectif affiché était alors de définir un premier jeu d'instructions généralistes manipulant des chiffres FHE pour une abstraction logicielle d'un processeur « 8 bits ».

Cette année, un jalon majeur a été franchi : l'équipe vient de proposer une solution reposant sur des techniques de débruitage (bootstrapping) programmable avancées et le cryptosystème TFHE. Un des enjeux majeurs est de s'efforcer de comprendre sa portée et d'en tirer le meilleur parti. Dans ce contexte, nos travaux explorent une piste (l'utilisation du bootstrapping de TFHE avec des chiffres d'entiers) encore peu étudiée pour le FHE généraliste, qui s'avère compétitive avec l'état de l'art

(qui, à l'instar de Cingulata, procède par évaluation de circuits Booléens sur des chiffres binaires).

Ainsi, l'équipe constituée autour de Daphné s'est donc attelée à l'élaboration d'un « processeur » homomorphe généraliste, dont l'ISA propose une cinquantaine d'instructions. Ce jeu d'instructions universel nous permet d'implémenter l'équivalent homomorphe de n'importe quel algorithme classique (tri de tableau, calcul de moyenne, trouver un maximum/minimum, et même, grande première, l'évaluation d'un neurone avec une fonction d'activation sigmoïde et des entrées/sorties à plusieurs décimales de précision, etc.). L'évaluation d'une demi-douzaine d'algorithmes à l'aide de notre jeu d'instruction permet de montrer que notre approche est en moyenne près de deux fois plus efficace que l'état de l'art. En particulier, notre approche se démarque dès que le circuit Booléen correspondant à l'instruction exécutée grandit (typiquement, une multiplication est suffisante).



DAPHNÉ TRAMA,
Doctorante



RENAUD SIRDEY
Directeur de recherche

Ces résultats sont résumés dans le tableau suivant, où les temps d'exécution sont donnés en secondes.

Algorithme	MadPanthera (List 2025)	Juliet	E3	Cingulata (List 2020)	Concrete
Maximum	4,74 s	7,91 s	3,13 s	1,87 s	21,52 s
BubbleSort	15,22 s	31,62 s	12,70 s	6,18 s	78,07 s
SquaresSum	5,66 s	25,10 s	8,82 s	8,40 s	7,69 s
Average	2,36 s		3,57 s	10,31 s	
Sigmoid	8,98 s			969,0 s	



Présentation de Daphné Trama à Kuala Lumpur lors de la prestigieuse conférence CHES 2025.

À noter, l'état de l'art le plus compétitif, Cingulata, reste significativement plus rapide que les solutions concurrentes.

Ces travaux identifient également les noyaux de calculs à optimiser/accélérer pour le FHE généraliste, ouvrant la voie à une collaboration avec des équipes de conception matérielle du List, voire à terme au développement d'un accélérateur dédié.

En lien avec ces travaux, nous avons travaillé sur une application emblématique et au cœur du sujet du premier appel du NIST (National Institute of Standards and Technology) visant du FHE : l'exécution homomorphe de l'AES (Advanced Encryption Standard). Après avoir amélioré l'état de l'art international d'un facteur 10 en 2023 [3], plusieurs équipes de différents pays ont successivement amélioré le « record » de l'exécution homomorphe de l'AES la plus rapide en proposant de nouvelles techniques. En 2025, par le biais d'une collaboration avec une équipe de CryptoExperts, la France a repris la tête de cette compétition amicale en proposant une nouvelle méthode plus efficace qui améliore le précédent record de 33%. Cette approche reprend les bases sous-jacentes de MadPanthera et l'étend avec une approche complémentaire dites des « *p-encodings* ».

50%

Réduction de moitié des temps d'exécution d'algorithmes de calcul homomorphe sur des chiffres

Projets associés

Ces travaux ont été réalisés dans le cadre des projets PEPR SecureCompute et TRUSTInclouds.

Publications majeures

Ces travaux ont été présentés lors de l'édition 2025 de la prestigieuse conférence CHES et font l'objet d'un article [1] dans le journal associé à la conférence (IACR Transactions on Cryptographic Hardware and Embedded Systems).



[1] « **Designing a General-Purpose 8-bit (T)FHE Processor Abstraction** ». Daphné Trama, Aymen Boudguiga, Pierre-Emmanuel Clet, Renaud Sirdey, Nicolas Ye: *IACR Trans. Cryptogr. Hardw. Embed. Syst.* 2025(2): 535-578 (2025)
<https://doi.org/10.46586/tches.v2025.i2.535-578>



[2] « **Further Improvements in AES Execution over TFHE** ». Sonia Belaïd, Nicolas Bon, Aymen Boudguiga, Renaud Sirdey, Daphné Trama, Nicolas Ye: *IACR Commun. Cryptol.* 2(1): 39 (2025)
<https://doi.org/10.62056/ahmp-4tw9>



[3] « **A Homomorphic AES Evaluation in Less than 30 Seconds by Means of TFHE** ». Daphné Trama, Pierre-Emmanuel Clet, Aymen Boudguiga, Renaud Sirdey: *WAHC@CCS 2023: 79-90*
<https://doi.org/10.1145/3605759.3625260>

Cybersécurité et numérique de confiance

Technologies de sécurisation

Accélérateur étroitement couplé à un cœur RISC-V pour la cryptographie post-quantique

La cryptographie post-quantique, conçue pour faire face à la menace que l'ordinateur quantique apporte à la cryptographie moderne, s'impose progressivement dans l'écosystème industriel. HQC, un des schémas de cryptographie post-quantique récemment standardisés par le NIST, présente des temps d'exécution relativement élevés sur les microcontrôleurs des systèmes embarqués. Pour relever ce défi, nous proposons TYRCA, un accélérateur pour HQC développé spécifiquement pour processeurs RISC-V.



STEFANO DI MATTEO
Ingénieur-chercheur



EMANUELE VALEA
Ingénieur-chercheur

EN PARTENARIAT AVEC
CEA-LETI ET POLITECNICO
DI TORINO

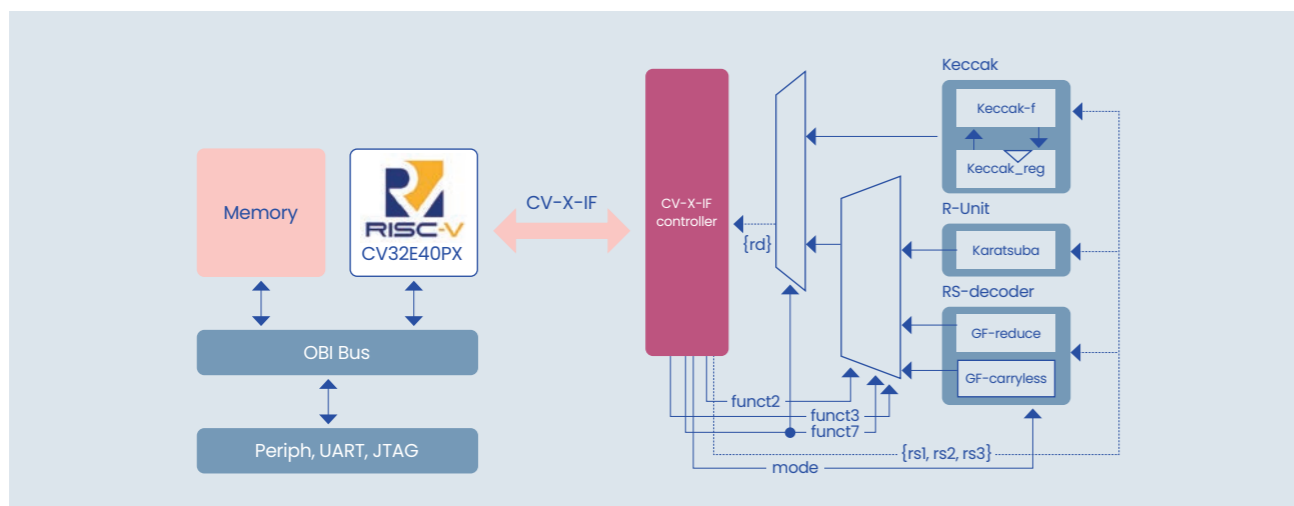


Schéma d'architecture d'un système sur puce avec le coprocesseur TYRCA couplé au cœur RISC-V CV32E40PX. Le coprocesseur est constitué d'un contrôleur CV-X-IF et d'un ensemble d'accélérateurs matériels spécifiques pour certains opérations d'HQC, notamment Keccak, la multiplication polynomiale (Karatsuba) et le décodage de Reed-Solomon (RS).

Les ordinateurs quantiques, susceptibles de révolutionner la médecine, les matériaux et l'intelligence artificielle, menacent la sécurité des communications. Les systèmes cryptographiques comme RSA, Diffie-Hellman et ECC sont

vulnérables aux attaques quantiques, ce qui motive le développement de la cryptographie post-quantique (PQC). HQC, un schéma fondé sur un formalisme mathématique dérivé de la théorie des codes et sélectionné

comme standard par le NIST en 2025, présente des performances relativement faibles par rapport aux autres candidats, basés sur des réseaux euclidiens. Bien que plus rapides, les implémentations matérielles sont coûteuses en surface silicium, la multiplication polynomiale sur de polynômes de très grande taille constituant le principal goulot d'étranglement.

Une alternative aux méthodes classiques d'accélération matérielle est l'accélération étroitement couplée, qui encapsule les opérations récurrentes de l'algorithme dans des instructions personnalisées ajoutées au jeu d'instruction du processeur. Grâce à l'interface Core-V extension (CV-X-IF), l'accélérateur s'intègre parfaitement dans le pipeline d'exécution des processeurs RISC-V, offrant un gain de vitesse transparent tout en surmontant les défis traditionnels d'extension de l'ISA.

Nous avons appliqué cette nouvelle stratégie d'accélération étroitement couplée et conçu trois blocs d'accélération matérielle dédiés ciblant les principaux goulots d'étranglement de l'algorithme HQC. La R-Unit accélère la multiplication polynomiale via une approche Karatsuba en plusieurs étapes sur des blocs de 32 bits avec quatre instructions personnalisées pour gérer complètement le résultat. Le RS-Decoder accélère les opérations clés sur les champs de Galois, dont la multiplication sans retenue et le comptage des zéros finaux, grâce à plusieurs instructions spécialisées éliminant boucles et résultats intermédiaires. Enfin, l'accélérateur Keccak gère l'état de permutation de 1600 bits avec un registre dédié et trois instructions personnalisées pour un chargement, un traitement et un stockage efficaces, minimisant les surcoûts par rapport aux approches classiques faiblement couplées. En appliquant cette stratégie d'accélération étroitement couplée, TYRCA montre des améliorations de performance par rapport à l'implémentation logicielle originale de HQC. À tous les niveaux de sécurité (HQC-128/192/256), les opérations de génération de clé (KeyGen), d'encapsulation et décapsulation (Encaps, Decaps) voient une réduction d'environ 95 % du nombre de cycles d'horloge. L'utilisation de la mémoire d'instructions est également fortement réduite. Implémenté sur une cible FPGA Kintex-7, TYRCA occupe moins de 26 % de la surface totale du système-sur-puce, le R-Unit fournissant le gain de performance le plus élevé tout en utilisant moins de 10 % de cette surface. Les métriques de vitesse normalisées (vitesse/surface) confirment que TYRCA surpasse les approches faiblement couplées précédentes.



Publication majeure

« TYRCA: A RISC-V Tightly-coupled accelerator for Code-based Cryptography ». Dolmeta, A., Di Matteo, S., Valea, E., Carmona, M., Loiseau, A., Martina, M., & Masera, G. (2025, March). In *2025 Design, Automation & Test in Europe Conference (DATE)* (pp. 1-7). <https://doi.org/10.23919/DATE64628.2025.10993202>
Cet article a été retenu comme "Best Paper Candidate" lors de la conférence DATE (12 articles sélectionnés sur 330 acceptés à la conférence et plus de 1200 soumissions). Il montre des accélérations d'un facteur 20 par rapport à l'implémentation logicielle de référence du schéma de cryptographie post-quantique HQC.

Figure 1 : Stratégie d'intégration dans le pipeline RISC-V avec CV-X-IF

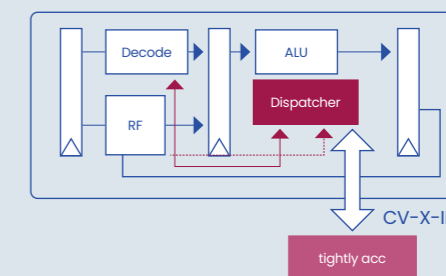


Figure 2 Décomposition de Karatsuba utilisée dans la R-unit

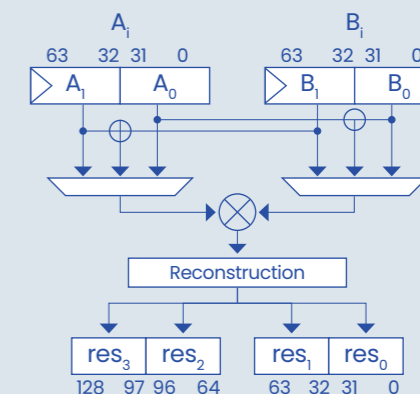


Figure 3 Réduction du nombre de cycles total sur HQC montrant les gains en performance

Version	Level	KGen+Enc+Dec
SW	128	407.90M
SW	192	1188.77M
SW	256	2182.38M
TYRCA	128	20.50M [-95%]
TYRCA	192	68.32M [-94%]
TYRCA	256	126.61M [-94%]

Retrouvez l'ensemble de nos résultats scientifiques lors des **List Days** Conference 2026 !

A U P R O G R A M M E

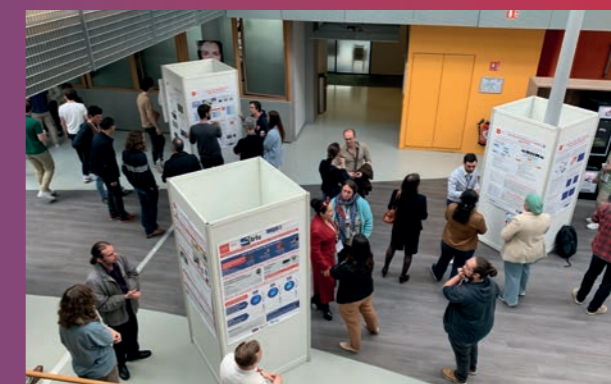
Des keynotes introductives

Des workshops thématiques autour de :

- Capteurs et instrumentation pour la mesure physique
- Calcul HPC & Quantique
- Ingénierie des données et des systèmes
- Collaboration Humain et systèmes intelligents
- Vers une IA plus réaliste
- Eco-innovation
- Cybersécurité
- IA pour les systèmes embarqués
- Vers une utilisation de l'IA plus fiable
- Modélisation physique, simulation des systèmes et composants critiques

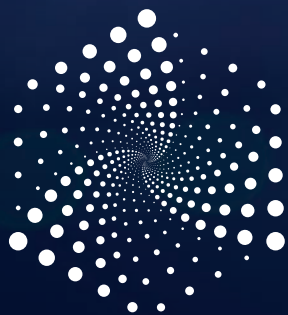


Deux sessions de rencontres
et d'échanges autour de posters
scientifiques présentant les résultats
des étudiants en thèse



10 & 11 juin 2026 – Nano-INNOV – Palaiseau
VOTRE RENDEZ-VOUS AU CŒUR DE LA SCIENCE





Excellence scientifique
et technologique



Créer, animer des écosystèmes
Nationaux et Européens



Répondre aux besoins
industriels

