

Avis de Soutenance



Monsieur Mathieu CHÊNE

Spécialité "Electronique, microélectronique, nanoélectronique et micro-ondes"

Soutiendra publiquement ses travaux de thèse intitulés

Circuit intégré d'intelligence artificielle pour capteurs intelligent ultra basse consommation

dirigés par Monsieur Antoine FRAPPÉ et Monsieur Andreas KAISER

Soutenance prévue le **mardi 16 décembre 2025** à 13h30

Lieu : 41 Boulevard Vauban, Bâtiment IC1, 59800 Lille, France

Salle : Amphi Mathilda

Composition du jury proposé

M. Antoine FRAPPÉ	Université de Lille	Directeur de thèse
M. Slavisa JOVANOVIC	Université de Lorraine	Rapporteur
M. François RIVET	Université de Bordeaux	Rapporteur
Mme Shih-Chii LIU	University of Zurich	Examinatrice
M. Andreas KAISER	Université de Lille	Co-directeur de thèse
M. Wim DEHEANE	KU Leuven	Examineur
M. Amirali AMIRSOLEIMANI	York University	Invité
M. Benoît LARRAS	Université de Lille	Invité

Mots-clés : Réseaux de capteurs, Intelligence artificielle, calcul intra-réseau, Entraînement sur puce,

Résumé :

Dans un contexte où l'edge computing prend une importance croissante, le passage des architectures classiques de type Von Neumann vers des architectures de calcul en mémoire (Compute-In-Memory, CIM) apparaît comme une solution prometteuse pour réduire la consommation énergétique des systèmes intelligents déployés dans des environnements contraints. Alors que la majorité des recherches se concentrent sur l'optimisation de l'inférence sur puce, des travaux plus récents explorent l'implémentation de l'apprentissage embarqué, notamment le calcul de l'erreur et des gradients, afin d'apporter davantage de flexibilité aux systèmes en périphérie. Néanmoins, la mise à jour des paramètres reste un défi majeur en matière d'efficacité énergétique, en raison des mouvements de données en mémoire intrinsèques à cette opération. Ce travail de thèse propose une macro numérique, réalisée en 28nmFDSOI ST Microelectronics, pour le calcul en mémoire, reposant sur une architecture d'inférence de type bit-série avec des entrées et des poids sur 8 bits, ainsi qu'une architecture de mise à jour des poids en mémoire. Cette dernière utilise des gradients encodés sous forme de puissances de deux, associés à une cellule SRAM spécialement conçue pour permettre une inversion locale de bits. Les résultats montrent une efficacité énergétique pouvant atteindre 58.9 TOPS/W pour l'inférence et jusqu'à 527 TOPS/W pour la mise à jour des poids. Ce travail permet de multiplier par un facteur 9 l'efficacité énergétique pour la mise à jour de poids par rapport à l'approche classique Lecture-Calcul-Ecriture et ce sans impacter l'efficacité énergétique de l'inférence. Ce travail de thèse, qui introduit le concept de Training-In-Memory (TIM), ouvre des perspectives vers des systèmes embarqués intelligents dont le cœur IA peut se mettre à jour en fonction du contexte et sans nécessiter l'usage de grosses infrastructures de type serveurs.