

## Simulation au niveau système et exploration d'architectures neuromorphiques non-volatiles

**Contact :** François DUHEM DRF//INAC/SPINTEC [Francois.DUHEM@cea.fr](mailto:Francois.DUHEM@cea.fr) 04 38 78 52 98

**Stage pouvant se poursuivre en thèse :** Non

### Résumé :

L'implantation matérielle de réseaux de neurones est un sujet de recherche stratégique pour de nombreuses entreprises internationales. Les principaux projets autour de l'ingénierie neuromorphique ont donné naissance à des puces inspirées du comportement du cerveau comme SyNAPSE, TrueNorth ou SpiNNaker. Ces technologies ciblent principalement de puissantes fermes de calcul et sont peu adaptées aux contraintes de consommation énergétique des systèmes embarqués ou de l'internet des objets.

L'intégration hétérogène de la technologie CMOS avec des technologies émergentes permettrait de s'affranchir de ces limitations. En particulier, la technologie mémoire MRAM (Magnetoresistive Random-Access Memory) est considérée comme la plus prometteuse des mémoires non-volatiles permettant de réduire la consommation énergétique des architectures de calcul. Il manque toutefois d'une approche haut niveau permettant d'évaluer et d'améliorer les gains apportés par ces mémoires. Dans le cadre d'une collaboration entre Spintec et le LEAT à Sophia Antipolis, un simulateur d'architecture de réseaux de neurones impulsionsnels a été développé en SystemC afin d'évaluer le coût en termes de surface et de consommation d'énergie ainsi que les performances de différentes architectures, permettant ainsi l'exploration de l'espace de conception des architectures neuromorphiques non-volatiles.

Dans ce contexte, ce sujet de stage consiste à continuer le développement de ce simulateur et de l'enrichir de nouvelles fonctionnalités. En particulier, le stage permettra de raffiner la modélisation des mémoires et de faire la comparaison avec des architectures existantes afin d'estimer la précision de la simulation. Une application dans le domaine de la vision sera utilisée pour démontrer les capacités du simulateur à évaluer les avantages d'une architecture non-volatile.

### Sujet détaillé :

L'implantation matérielle de réseaux de neurones est un sujet de recherche stratégique pour de nombreuses entreprises internationales. L'intérêt grandissant pour la reconnaissance de motifs à l'aide de réseaux de neurones profonds a principalement bénéficié à de grandes entreprises américaines comme NVidia, Google, IBM ou Intel. Leur approche repose majoritairement sur des GPGPU (General-Purpose Graphics Processing Unit) pour l'apprentissage et sur des puces dédiés pour l'inférence sur des cibles embarquées à des fins d'efficacité énergétique. Les principaux projets autour de l'ingénierie neuromorphique ont donné naissance à des puces inspirées du comportement du cerveau afin de développer de nouveaux paradigmes de calcul (SyNAPSE, TrueNorth) ou de contribuer à la recherche en neuroscience à travers le Human Brain Project (SpiNNaker). Ces technologies ciblent principalement de puissantes fermes de calcul et sont peu adaptées aux contraintes de consommation énergétique des systèmes embarqués ou de l'internet des objets.

L'intégration hétérogène de la technologie CMOS avec des technologies émergentes permettrait de s'affranchir de ces limitations. En particulier, la technologie mémoire MRAM (Magnetoresistive Random-Access Memory) est considérée comme la plus prometteuse des mémoires non-volatiles. Grâce à des performances proches de mémoires SRAM et une bonne compatibilité avec l'évolution des n?uds technologiques, la MRAM permettrait de réduire la consommation énergétique des architectures de calcul embarquées.

En particulier, la MRAM pourrait être utilisée dans les processeurs non-volatils qui intègrent des éléments de stockage non-volatils dans la hiérarchie mémoire afin de réduire la consommation énergétique. Plusieurs études ont mis en avant le compromis entre la performance, la taille et la consommation énergétique de telles architectures hybrides pour des processeurs généralistes, mais elles ne considèrent pas les accélérateurs

spécifiques à un domaine d'application qui sont souvent nécessaires à la réalisation de meilleures performances. Dans ce contexte, une collaboration entre Spintec et le LEAT (Laboratoire d'Electronique, Antennes et Télécommunications) à Sophia Antipolis a permis de développer un simulateur SystemC permettant l'évaluation d'architectures de réseaux de neurones impulsionsnels et fournissant des métriques sur l'implantation comme la performance, la surface utilisée ou la consommation d'énergie. Le simulateur doit ainsi permettre l'adoption d'architectures neuromorphiques à base de mémoires non-volatiles pour des domaines d'application allant des objets connectés à la robotique ou aux véhicules autonomes.

Ce sujet de stage consiste à continuer les développements sur le simulateur tout en y ajoutant de nouvelles fonctionnalités qui permettraient une meilleure comparaison entre les architectures volatiles et non-volatiles. Les missions du stage seront les suivantes :

- Etude bibliographique sur les mémoires et architectures non-volatiles
- Raffinement des modèles de mémoire
- Développement de nouvelles fonctionnalités du simulateur
- Evaluation de la précision du simulateur
- Démonstration sur une application à définir dans le domaine de la vision

**Compétences requises :**

Connaissance en architecture des systèmes embarqués, électronique et en programmation C/C++. Des compétences en SystemC et en développement RTL sont un plus.