

ULPAI

"Feature Neural Networks : a new paradigm for ultra low-power AI"

Partenaires :

- Laboratoires Université de Toulon :
 - Porteur : IM2NP (UMR 7334)– Encadrement : Valentin GIES
- Partenaire socio-économique :
 - Société HIPPY : spécialisées dans les airbags pour personnes âgées.

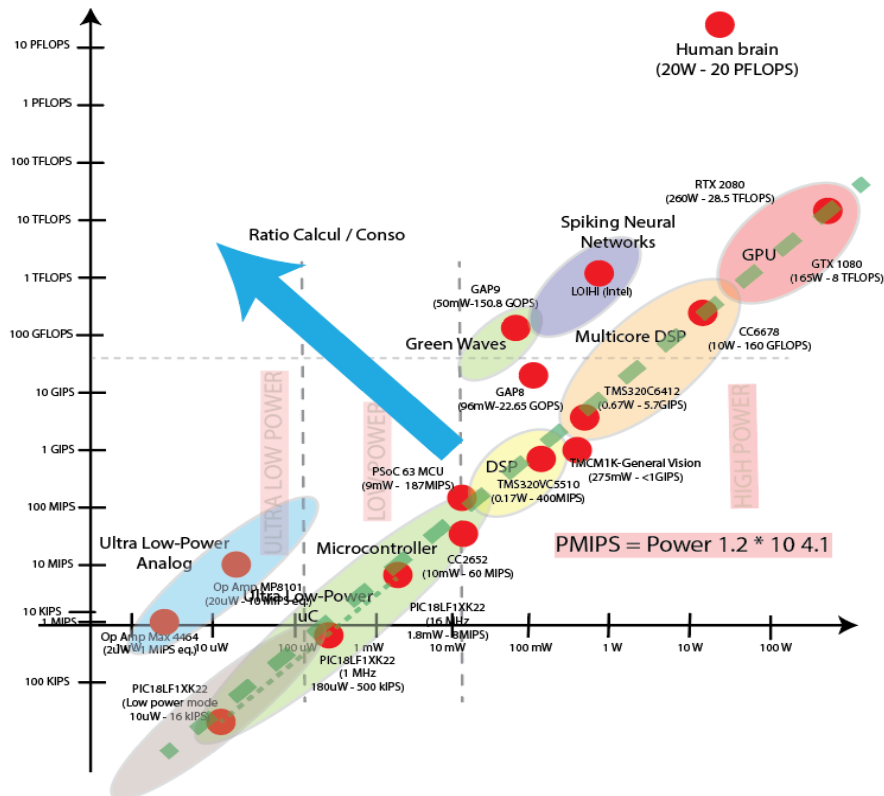
Contexte de la thèse

La consommation électrique actuelle liée aux technologies numériques dans le monde représente aujourd'hui 7.7% de la consommation mondiale. Elle passera à 20% en 2025, puis à 30% en 2030 (*Source : Total Consumer Power Consumption Forecast –Anders S.G. –Nordic Digital Business Summit*).

Dans le même temps, l'intelligence artificielle entre dans tous les domaines de la société, avec une consommation énergétique toujours plus importante, en particulier en raison de la décentralisation des traitements de données vers des serveurs distants.

L'intelligence artificielle à ultra faible consommation d'énergie est donc un enjeu majeur pour que l'intelligence artificielle puisse s'intégrer de manière soutenable dans les décennies à venir.

L'IM2NP en collaboration avec le LIS travaille depuis plusieurs années sur cette thématique, avec un objectif : repousser les limites des architectures classiques en terme de ratio puissance de calcul / consommation énergétique, avec un focus sur les architectures à très faibles ressources. Partant du constat que la relation entre puissance de calcul et consommation énergétique est linéaire en échelle logarithmique dans les architectures classiques, cela suppose un véritable changement de paradigme en matière d'adéquation algorithme architecture dans le domaine de l'IA. En effet, le parallélisme et l'utilisation de circuits MAC dédiés (DSP) permettent de gagner en puissance de calcul, au prix d'une augmentation quasi proportionnelle de la consommation énergétique dans les architectures implantant l'intelligence artificielle de manière classique (microcontrôleurs, DSP, GPU), ce qui n'apporte aucune rupture en termes de consommation énergétique.



Sortir de ce paradigme n'est pas impossible : nous avons exploré dans la thèse (2018-2021) de doctorat de SebastianMarzetti comment réduire drastiquement la consommation d'énergie à puissance de calcul constante dans le domaine de l'intelligence artificielle ultra-low power (<100uW). Un tour d'horizon des algorithmes d'apprentissage supervisés ou non supervisés sur architectures à faible ressources a démontré qu'une des pistes les plus prometteuses pour obtenir un fonctionnement ultra-low power est d'utiliser des primitives analogiques. L'utilisation de primitives analogiques afin de prétraiter les signaux haute fréquence permet de gagner deux ordres de grandeur de consommation énergétique par rapport à l'utilisation de traitement numériques classiques.

Le but de la thèse proposée est de poursuivre ces travaux en se focalisant sur la généralisation de ces résultats à l'ensemble de l'IA low-power en proposant un nouveau paradigme algorithme-architecture pour l'IA ultra low-power et une nouvelle famille de réseau d'apprentissage que nous appelons les Feature Neural Networks (FNN). Ils reposent sur l'usage de primitives efficaces ne pouvant être implantées ou optimisées facilement à l'aide de réseaux type CNN, en particulier en raison de leur mémoire temporelle. Ces réseaux sont composés, à l'instar du « ReservoirComputing », de primitives à topologie figée (mais pouvant être ajustées via un jeu de paramètres), suivi d'une couche fortement connectée de neurones standards. L'utilisation de ces primitives (les featureextractors) en analogique permet de diviser le coût énergétique de prétraitement par 100. Cette approche parcimonieuse permet donc de gagner fortement (environ deux ordres de grandeur) en consommation d'énergie par rapport aux techniques classiques.

Cette approche pose toutefois un certain nombre de difficultés : tout d'abord, elle nécessite de disposer d'une architecture silicium pour son implantation. Celle-ci sera au cœur de la thèse, avec la réalisation d'un chip silicium dédié (ASIC). D'un point de vue algorithmique,

les problématiques théoriques de convergence de l'apprentissage sur ce type de réseau seront également au cœur de cette thèse. En particulier l'usage des primitives analogiques permet d'effectuer l'extraction de moments sur des signaux temporels, réduisant de manière drastique la bande passante du signal pertinent. Ces primitives, particulièrement pertinentes d'un point de vue algorithmique et efficacité énergétique, peuvent en revanche poser des difficultés lors de l'apprentissage : la thèse aura pour objectif de les résoudre.

Sujet de thèse

L'intitulé du sujet de thèse est : **"Réseaux Neuronaux de Primitives : un nouveau paradigme pour l'Intelligence Artificielle Ultra-Low Power"**, ou en version anglaise : **"Feature Neural Networks : a new paradigm for ultra low-power AI"**.

La thèse s'inscrit dans la suite de celle de Sebastian Marzetti, qui se terminera en Septembre 2021. Le candidat devra reprendre les travaux déjà engagés, et en particulier ceux sur les Feature Neural Network, couple architecture/algorithmes destiné à utiliser de manière optimale les technologies analogiques et numériques afin de minimiser la dépense énergétique liée aux algorithmes d'intelligence artificielle.

La réalisation d'un circuit intégré dédié à l'implantation des Feature Neural Network sera une des attentes de ce doctorat, en collaboration avec l'ensemble de l'équipe CCIS de l'IM2NP. Ce circuit pourra faire l'objet d'une valorisation en termes de propriété intellectuelle sous forme de brevet par exemple.

Des applications en IA ultra-low power seront également proposées, dans un premier temps sur des architectures classiques et dans un second temps sur l'architecture dédiée aux Feature Neural Networks. En particulier, des applications dans le domaine de la santé et de la SilverEconomy seront proposées en coopération avec la société Hippy.

Du point de vue du partenaire socio-économique

HIPPY est une start-up de dix personnes spécialisée dans l'intelligence artificielle et la miniaturisation de technologies dans le domaine de la Prévention Santé des personnes âgées.

Le concept du bien vieillir est à l'origine des préoccupations d'Hippy.

Le projet « Ceinture » est né pour lutter contre le « fléau » que représentent les fractures du col du fémur. Il n'existe aucune solution préventive efficace et 25% des patients de plus de 55 ans décèdent des suites de l'opération. Les dépenses représentent globalement 6 milliards d'euros par an.

L'activité est organisée autour des compétences nécessaires au développement d'un système de ceinture qui permettra d'éviter la fracture du col du fémur en cas de chute. Elle intègre une forte dimension technologie et scientifique (Intelligence artificielle, Electronique

embarque, Système airbag, Miniaturisation de composants) mais également médicale, qualité et réglementaire.

La société Hippy s'appuie sur un comité scientifique composé de médecins et chercheurs du monde hospitalo-universitaire.

Le projet ceinture airbag initié par Hippy s'appuie sur deux années d'études collaboratives avec le numéro un mondial de fabrication d'airbags.

Prévision du déroulement

La première année sera consacrée d'abord à l'intégration de l'existant et à une remise à jour de l'état de l'art. Elle aboutira à une publication de synthèse en fin d'année dans un congrès international, portant sur l'ensemble des solutions technologiques permettant d'implanter de l'intelligence en ultra-low power dans les systèmes embarqués, en les comparant à la solution proposée des features neural networks, qui sera développée en profondeur en particulier sur un plan théorique au niveau des mécanismes d'apprentissage. Les algorithmes de détection de chute en ultra low-power nécessaires à la société HIPPY seront également développés dans une version embarquée sur microcontrôleur dans cette première phase.

Un jalon à 12 mois permettra de faire le bilan.

En deuxième année et après cet état de l'art, le focus sera mis sur le développement silicium d'un chip mixte analogique-numérique permettant d'implanter les Features Neural Networks.

La troisième année de travail permettra de commencer à utiliser le chip développé de manière à implanter efficacement divers algorithmes d'intelligence artificielle dans des cas d'usage concret en les comparant aux autres solutions existantes. En particulier une version améliorée des algorithmes permettant la détection de chute en ultra-low power et basées sur le processeur dédié aux Features Neural Network sera proposée. La rédaction du mémoire de thèse, la participation à un congrès international et la production d'un ou deux articles dans des journaux reconnus en électronique ou intelligence artificielle sont également prévues pour cette 3^e année.