

## Présentation du projet de recherche doctorale

### Impact du codage impulsionnel sur l'efficacité énergétique des architectures neuromorphiques

#### Contexte scientifique

Le projet européen Human Brain Project (HBP) ainsi que d'autres projets internationaux de grande ampleur ont mis en évidence que les recherches sur le cerveau numérique représenteraient un enjeu scientifique, voire stratégique, majeur du XXI<sup>e</sup> siècle. Si l'accent est mis dans un premier temps sur l'importance des simulations à grande-échelle pour la compréhension du fonctionnement et de la structure du cerveau [**HBP2015**], ces simulations doivent aussi permettre de mieux appréhender les processus cognitifs et de développer à terme de nouveaux algorithmes d'Intelligence Artificielle (IA), les réseaux de neurones profonds (Deep Networks) en étant l'exemple le plus récent.

Le projet de thèse proposé se positionne dans ce contexte scientifique en s'intéressant aux architectures électroniques neuromorphiques. Ces architectures s'inspirent des capacités de traitement de l'information du cerveau pour explorer de nouveaux paradigmes de calcul comparés aux processeurs conventionnels, dits de Von Neumann.

De nombreuses architectures matérielles ont été proposées dans la littérature internationale.

Elles prennent la forme soit d'architectures neuromorphiques électroniques ([**TrueNorth2014**] en collaboration entre IBM et l'Université de Cornell, [**BrainScales2014**] à l'Université d'Heidelberg,...) soit de calculateurs plus traditionnels adaptés à ce modèle de calcul neuronal ([**Spinnaker2016**] à l'Université de Manchester, architecture GPU JetsonTX1 de NVidia, ...).

Ces architectures font le pari, soit de se rapprocher de la structure et du comportement des neurones biologiques (pour les systèmes neuromorphiques), soit d'utiliser une structure conventionnelle (en misant sur la puissance des processeurs plutôt que sur leur mimétisme structurel avec les neurones). Si ces deux types d'architectures deviennent indispensables pour la simulation à grande-échelle de réseaux de neurones, elles participeront aussi au développement des architectures matérielles du futur (meilleure scalabilité, calcul distribué, capacités d'apprentissage...) et faciliteront la diffusion d'algorithmes d'IA et de machine learning dans différents secteurs industriels.

#### Objectifs du projet de recherche

L'efficacité énergétique des architectures neuromorphiques reste encore un sujet mal étudié du fait de la nouveauté de ce problème interdisciplinaire et de la dépendance à des facteurs technologiques très variables. La mesure et l'évaluation de la consommation des architectures neuromorphiques numériques comparées aux architectures standards de Von Neumann (processeurs) constitue donc le cadre général de cette thèse.

Les algorithmes neuronaux considérés aujourd'hui à la fois dans la littérature scientifique et dans les utilisations industrielles proviennent essentiellement de deux domaines encore dissociés : le machine learning (dans lequel sont par exemple étudiés les réseaux de neurones profonds – deep learning - massivement utilisés pour l'analyse de données ou data sciences) et les neurosciences (qui cherchent à modéliser de manière plus réaliste le fonctionnement du cerveau). Les modèles de neurones artificiels de ces deux domaines exhibent des caractéristiques très différentes et ce sujet s'intéressera à rapprocher les deux domaines scientifiques en déclinant le codage impulsionnel (à spikes [**LIF08**]) de l'information neuronale provenant des neurosciences dans celui du machine learning. Cette thèse défendant l'idée que

c'est justement ce codage bio-inspiré de l'information qui rend le cerveau biologique aussi efficace d'un point de vue énergétique.

Cependant la question du codage de l'information au niveau biologique fait encore débat dans la communauté des neurosciences. Le codage par rang [**Thorpe01**], le codage temporel [**Brette2015**] ou encore le codage par population présent dans les modèles DNF [**Amari77**] restent à explorer et à comparer du point de vue matériel.

Une des contributions importantes proposée par ce projet est donc de fournir une étude complète et quantitative du coût énergétique de tel ou tel type de codage neuronal dans les circuits électroniques numériques [**NPU2015**]. Les solutions proposées seront comparées à une exécution logicielle optimisée sur la plateforme Spinnaker disponible dans le projet NeuComp de l'Université Côte d'Azur [**NeuComp**]. Les seuls résultats quantitatifs publiés sur le sujet datent de décembre 2015 et n'ont été effectués que sur de petites tailles de réseaux [**Micro2015**]. Il convient donc ici de prolonger l'étude sur des réseaux de taille réaliste et de mieux évaluer l'impact du codage impulsif sur la consommation des accélérateurs neuromorphiques qui, à n'en pas douter, seront prochainement intégrés dans de nombreux domaines applicatifs, des smartphones jusqu'aux véhicules autonomes. Nous proposons dans ce projet de nous appuyer sur la plateforme conception CimPaca (<http://www.pf-conception.org/>), et les outils de CAO micro-électronique associés, pour développer les premiers prototypes de blocs IP qui pourraient à terme être intégrés dans des solutions SoC (system-on-chip) industrielles.

## Hardware design of spiking neural networks for energy-efficient brain inspired computing

### Scientific context

The European Human Brain Project (HBP) and other large-scale international projects have shown that research on the digital brain would represent a major scientific, and even strategic, challenge of the 21st century. If the emphasis is firstly on the importance of large-scale simulations for understanding the functioning and structure of the brain [HBP2015], these simulations should also help to better understand the cognitive processes and to finally develop new Artificial Intelligence (AI) algorithms, deep neural networks being the most recent example. The proposed doctoral project is positioned in this scientific context by focusing on neuromorphic electronic architectures. These architectures are inspired by the information processing capabilities of the brain to explore new computational paradigms compared to conventional Von Neumann processors.

Numerous hardware architectures have been proposed in the international literature. They take the form either of electronic neuromorphic architectures ([TrueNorth2014] in collaboration between IBM and the University of Cornell, [BrainScales2014] at the University of Heidelberg, ...) or of more traditional computers adapted to this model of computation ([Spinnaker2016] at the University of Manchester, NVidia GPU architecture JetsonTX1, ...). These architectures try either to mimic the structure and behavior of biological neurons (for neuromorphic systems), or to use a more conventional structure (relying on the power of processors rather than on their structural mimicry with neurons). If these two types of architecture become indispensable for the large-scale simulation of neural networks, they will also participate in the development of future hardware architectures (better scalability, distributed computing, learning abilities ...), or embedded versions of machine learning algorithms in different industrial sectors.

### Contributions of the thesis

The energy efficiency of neuromorphic architectures is still poorly studied because of the novelty of this interdisciplinary problem and its dependence on widely varying technological factors. Measurement and evaluation of the consumption of digital neuromorphic architectures compared to standard Von Neumann architectures (processors) is therefore the general framework of this thesis.

The neuronal algorithms considered today both in the scientific literature and in industrial applications are essentially derived from two still dissociated domains: machine learning (in which, for example, deep neuron networks - deep learning – are massively used for analysis of data) and neurosciences (which seek to model more realistically the functioning of the brain). The artificial neural models of these two domains exhibit very different characteristics and this subject will be interested in bringing the two scientific domains closer together by declining the impulse coding (to spikes [LIF08]) of the neuronal information coming from the neurosciences in that of the machine learning. This thesis defends the idea that it is precisely this bio-inspired coding of information that makes the biological brain so efficient from an energy point of view. However, the issue of information coding at the biological level is still debated in the neuroscience community. The coding by rank [Thorpe01], the temporal coding [Brette2015] or the population coding present in the DNF models [Amari77] remain to be explored and compared from the hardware point of view.

One of the important contributions proposed by this project is therefore to provide a complete and quantitative study of the energy cost of these types of neural coding in digital electronic circuits [NPU2015]. The proposed solutions will be compared to the software execution onto the Spinnaker platform available in the NeuComp Project of the Université Côte d'Azur

**[NeuComp]**. The only quantitative results published on the subject have been carried out only on small network sizes **[Micro2015]**. It is therefore appropriate here to extend the study on networks of realistic size and to better evaluate the impact of impulse coding on the consumption of neuromorphic accelerators which will undoubtedly soon be integrated into many application domains, such as smartphones to autonomous vehicles. We propose in this project to build on the design platform CimPaca (<http://www.pf-conception.org/>), and the associated microelectronic CAD tools, to develop the first prototypes of IP blocks that could ultimately be integrated into industrial SoC (system-on-chip) solutions.

**Candidatures à envoyer avant le 31 aout :**

Envoyer un CV, une lettre de motivation, un relevé de notes de la dernière année de formation ainsi que d'éventuelles lettres de recommandation.

**Application before 31th august:**

Send a CV, a letter of motivation, a transcript of the last year of training and any letters of recommendation.

<b>Sigle :</b>	LEAT
<b>Nom Complet :</b>	Laboratoire d'Electronique, Antennes et Telecommunications UMR 7248 CNRS Université Côte d'Azur
<b>Adresse :</b>	Campus SophiaTech - Bâtiment Forum 930 route des Colles, BP 145 06903 Sophia Antipolis cedex
<b>Directeur :</b>	Jean-Yves Dauvignac
<b>Directeur de thèse :</b>	Benoît Miramond - Professor
<b>Contact :</b>	<a href="mailto:benoit.miramond@unice.fr">benoit.miramond@unice.fr</a>
<b>Téléphone :</b>	04.92.94.28.84

## BIBLIOGRAPHIE

[Amari77] Amari S: Dynamics of pattern formation in lateral inhibition type neural fields. *Biological Cybernetics*. 1977, 27: 77-87.

[BioComp] GDR BioComp, <http://gdr-biocomp.fr/>.

[BrainScales2014] Physical Models of Neural Circuits in BrainScaleS and the Human Brain Project Status and Plans, K. Meier, Neuro-Inspired Computational Elements Workshop, February, 2014

[Brette2015] Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain, R. Brette, *frontiers in Systems Neuroscience*, 2015

[HPB2015] Egidio D'Angelo et al., The Human Brain Project: High Performance Computing for Brain Cells Hw/Sw Simulation and Understanding, Euromicro Conference on Digital System Design (DSD), 2015.

[LIF08] W. Mass, Networks of spiking neurons: the third generation of neural network models, *Neural Network* 10 (9), pp. 1659-1671, 2008

[Micro2015] Z. DU, et. Al, Neuromorphic Accelerators: A Comparison Between Neuroscience and Machine-Learning Approaches, Proceedings of the 48th International Symposium on Microarchitecture, Pages 494-507, December, 2015

[NeuComp] NeuComp Project, IDEX UCA Jedi, <http://univ-cotedazur.fr/english/idx-uca-jedi/academies/networks-information-and-digital-society/recherche/liste-des-projets-finances/neucomp>

[NPU2015] Laurent Fiack, Benoit Miramond, Laurent Rodriguez, Hardware design of a neural processing unit for bio-inspired computing, IEEE 13th International on New Circuits and Systems Conference (NEWCAS), 2015.

[Spinnaker2016] Knight, James C. and Furber, Steve B. Synapse-Centric Mapping of Cortical Models to the SpiNNaker Neuromorphic Architecture, *Frontiers in Neuroscience*, vol.10, p.420, 14 September 2016.

[Thorpe01] Spike-based strategies for rapid processing, S. Thorpe et al, *NN*, 2001

[TrueNorth2014] Paul A. Merolla et al., A million spiking-neuron integrated circuit with a scalable communication network and interface, *Science*, 2014.