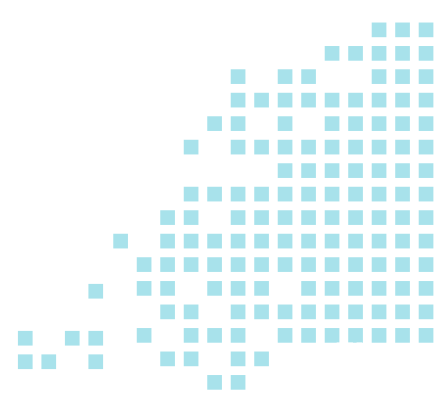




Laboratoire
d'Informatique
Fondamentale
de Lille



Sujet de thèse : modèle impulsif de réseaux de neurones artificiels à base de memristors, codage et architecture

Encadrants : Pierre Boulet et Philippe Devienne

15 avril 2013

Ce sujet est financé sur la période septembre 2013 à août 2016 par un contrat doctoral de l'université Lille 1 cofinancé par la région Nord Pas-de-Calais.

Ce sujet se place dans le contexte des nouveaux paradigmes de traitement de l'information, et plus précisément des accélérateurs neuromorphiques.

1. Contexte scientifique

La limite à la consommation d'énergie des puces électroniques (130 W, atteinte en 2004), la fin de la scalabilité de Dennard (qui permet de diminuer la consommation d'énergie d'un même facteur lors de la miniaturisation des transistors) et la plus grande variabilité des composants lorsque leurs dimensions diminuent posent des problèmes qui pourraient bien être insurmontables à la loi de Moore. Nous approchons ainsi de la fin d'un cycle de croissance exponentielle de la puissance de calcul des ordinateurs sur laquelle a reposé une grande partie des progrès du traitement de l'information ces 50 dernières années et pour partie aussi la croissance économique mondiale. Les projections actuelles disent qu'au mieux nous pourrions graver des puces en environ 10 nm au lieu des 20 nm actuels. Il ne resterait donc au rythme actuel que 4 à 6 ans de miniaturisation avec la technologie CMOS qui domine très largement le marché des puces électroniques depuis au moins 20 ans.

D'autre part, un nouveau composant électronique, le memristor, a été démontré en 2008 dans les laboratoires de Hewlett-Packard. Ce composant, dont plusieurs réalisations existent aujourd'hui a un comportement proche de celui des synapses dans les réseaux de neurones biologiques. Il est alors tentant d'imaginer réaliser des

réseaux de neurones artificiels à base de memristors. Le cœur du sujet proposé est d'explorer les architectures d'accélérateurs matériels à base de memristors ainsi que les problématiques du codage de l'information et de l'apprentissage dans de telles architectures. Celles-ci sont particulièrement prometteuses car elles sont naturellement résistantes à la variabilité des composants et applicables à un grand nombre d'applications (voir par exemple la suite de benchmarks RMS proposée par Intel).

Plusieurs importants programmes de recherche internationaux ont été lancés sur ce nouveau modèle de calcul dit *neuromorphique* ou *neuro-inspiré*. On peut citer les projets de recherche européens *Human Brain* [1] pour une modélisation la plus précise possible du cerveau et *BrainScaleS* [2] qui vise à construire des systèmes électroniques analogiques dont l'architecture et les règles de calcul sont directement inspirées du cortex, le projet américain financé par DARPA, *Mometa, A Mind Made from Memristors*, le projet *Variability Expedition* financé par la NSF pour des composants plus robustes et moins gourmands en énergie. Ces chercheurs issus de la biologie, la biophysique, de la nanoélectronique et de la théorie du calcul essaient de construire des plates-formes expérimentales capables de reproduire la biophysique des systèmes neuronaux biologiques. La taille et le nombre de ses programmes de recherche sont à la hauteur des enjeux scientifiques et des retombées économiques attendues.

2. Caractères émergents et interdisciplinaires du sujet

Le sujet de recherche des architectures neuromorphiques à base de memristors est un sujet émergent au niveau international (les memristors ne sont apparus qu'il y a 5 ans). Ce sujet se situe à la convergence de trois axes très dynamiques de recherche et développement :

- en biologie par notre connaissance de plus en plus précise du fonctionnement du cerveau à travers les neurones, leurs interconnexions (synapses) et les modèles d'apprentissage par impulsions (STDP : Spike-Timing-Dependent Plasticity) ;
- en électronique par la découverte à l'échelle nanométrique de l'élément manquant, le memristor et la question de la fabrication d'architectures utilisant ce nouveau composant ;
- en informatique par l'intérêt que représentent les réseaux de neurones artificiels pour les applications qui tolèrent des résultats approximatifs comme par exemple pour la reconnaissance, la classification et l'indexation de données.

Ces trois approches, biologique, électronique et informatique, convergent aussi sur leur souci de prendre en compte la variabilité croissante des processus avec défaut, ce qui correspond bien aux capacités intrinsèques de robustesse des réseaux

de neurones.

Ce sujet est encore très récent dans l'équipe DART, groupe Émeraude. Nous avons commencé de premières recherches sur ce sujet très prometteur en 2012 grâce à un financement de l'IRCICA. Le point particulier sur lequel nous allons travailler est les architectures à base des memristors conçus par nos collègues de l'IEMN (Philippe Pernod et Nicolas Tiercelin d'une part, et Dominique Vuillaume et Fabien Alibart d'autre part). Cette proximité géographique et le soutien de l'institut pluridisciplinaire IRCICA sont des gages d'une collaboration féconde.

3. Programme de travail

Une première simulation, tant logicielle que matérielle, a été réalisée par Olivier Bichler. Il a démontré dans sa thèse [3] que les architectures neuromorphiques sont bien adaptées pour l'implémentation d'algorithmes d'apprentissage non supervisés, issus des neurosciences cognitives, comme le modèle impulsionnel (STDP), ne nécessitant que peu de circuits de contrôle.

Le programme de cette thèse est de poursuivre dans cette voie et d'expérimenter d'autres applications dans le domaine de la reconnaissance vidéo ou audio sur des simulateurs de réseaux de neurones, et en collaboration avec les chercheurs de l'IEMN d'implémenter ces nouveaux algorithmes sur des matrices de type crossbar de plusieurs dizaines de milliers de memristors. Une thèse « Composants Memristifs pour approches neuromorphiques » est proposée par Philippe Pernod et Nicolas Tiercelin de l'IEMN. Une telle réalisation permettrait des applications réalistes avec des gains en temps de calcul et en consommation énergétique très significatifs.

Ce travail de thèse consistera donc en

- l'étude des simulateurs logiciels et matériels de réseaux de neurones à impulsions ;
- l'expérimentation de ces simulateurs sur des applications, en particulier dans le domaine de la vision, et de l'apprentissage supervisé et non supervisé, un travail théorique portera sur le codage de l'information sous forme impulsionnelle ;
- l'implémentation des réseaux de neurones sur des architectures à base de memristors, celles en cours d'élaboration à l'IEMN.

4. Retombées scientifiques et économiques potentielles

Les algorithmes de vision par ordinateur et leur mise en œuvre en temps réel via des accélérateurs neuromorphiques constituent un champ d'application très riche,

par exemple pour la navigation de véhicules autonomes, les systèmes de caméras intelligentes, la surveillance et la réalité augmentée.

Si de tels accélérateurs deviennent une réalité, ils pourraient apporter un gain de performance temporelle et énergétique spectaculaire (de plusieurs ordres de grandeur) pour les parties d'applications qu'ils exécuteraient. Or, ces applications sont de plus en plus importantes : classification, apprentissage machine, traitement approximatif de grandes quantités de données font en effet partie des grands défis de l'informatique d'aujourd'hui et de demain.

5. Collaborations prévues

Cette thèse s'inscrit dans une collaboration que nous menons depuis plus d'un an avec plusieurs équipes de recherche de l'EMN, tout particulièrement avec Philippe Pernod (professeur Centrale Lille) et Nicolas Tiercelin (chargé de recherche CNRS) qui souhaitent démarrer une thèse à l'EMN pour la réalisation d'une matrice 2D type crossbar pouvant connecter deux couches de 256 neurones avec 65536 (256^2) synapses.

Les financements et actions menées ou en cours en liaison avec cette demande de thèse sont :

- avec Philippe Pernod et Nicolas Tiercelin nous avons obtenu un projet BQR Lille 1 émergent « Composants Memristifs et approches neuromorphiques » ;
- avec Dominique Vuillaume (directeur de recherche CNRS) et Alexis Vlandas (chargé de recherche CNRS) nous avons organisé un workshop sur les architectures et le calcul bio-inspirés en décembre 2012 à l'IRCICA (http://www.lifl.fr/emeraude/?page_id=175, ces conférences étant accessibles sur Lille1 TV).
- nous collaborons au sein de l'IRCICA avec ces équipes dans le cadre du projet « Interfaces entre informatique et architectures bio-inspirées » financé par l'IRCICA en 2012 et 2013 au titre des actions de la Mission pour l'Interdisciplinarité du CNRS ;
- nous avons par ailleurs participé ensemble au workshop MEMCO [4] qui réunissait les meilleurs spécialistes de ce domaine.

Références

- [1] *Human Brain Project*, EU Flagship project, Janvier 2013, <http://www.humanbrainproject.eu>.
- [2] *The 6 Key Concepts of BrainScaleS*, Janvier 2011, <http://brainscales.kip.uni-heidelberg.de>.

- [3] Olivier Bichler, *Contribution à la conception d'architecture de calcul auto-adaptative intégrant des nanocomposants neuromorphiques et applications potentielles*, Université de Paris Sud, 14 Novembre 2012.
- [4] MEMCO (Memristors for Computing), Novembre 2012, Fréjus, France, <http://www.trt.thalesgroup.com/ump-cnrs-thales/memco>.