

---

# APPRENTISSAGE PAR RENFORCEMENT ET NEURONES IMPULSIONNELS : DES MODÈLES AU MATÉRIEL

équipe BISCUIT, Loria

*Encadrement* : Alain DUTECH (HDR), Bernard GIRAU (HDR)  
[alain.dutech@loria.fr](mailto:alain.dutech@loria.fr) & [bernard.girau@loria.fr](mailto:bernard.girau@loria.fr)

---

dernière modification le 6 mai 2019

## 1 Contexte

L'équipe BISCUIT<sup>1</sup>, est une équipe du laboratoire Loria<sup>2</sup> qui étudie de nouveaux paradigmes computationnels où les calculs sont adaptatifs, distribués et décentralisés, réalisés par des populations d'unités de calcul simples qui communiquent principalement avec leurs proches voisins. Ces propriétés sont compatibles avec la mise en œuvre de principes d'auto-organisation non-supervisés, mais guidés, pour s'attaquer à des problèmes difficiles comme le calcul cognitif situé, la robotique autonome, l'allocation adaptative de ressources de calcul, etc.

Ces caractéristiques permettent également d'envisager une meilleure utilisation des processeurs dit "neuromorphiques" qui sont en train d'émerger (IBM Truenorth, Intel Loihi, etc.). Ces processeurs sont basés sur des principes neuro-inspirés qui respectent les contraintes des paradigmes que nous étudions, et peuvent bénéficier des mécanismes d'auto-organisation non supervisée et guidée que nous mettons au point, à la fois en termes d'applications et en termes de gestion des ressources neuromorphiques. C'est pourquoi l'équipe BISCUIT s'attache à concevoir des architectures et des algorithmes d'apprentissage non supervisés et guidés pour des populations de calcul spatialisées et décentralisées tout en restant au plus près des contraintes et caractéristiques matérielles. Le sujet de la thèse de doctorat proposée est un pas de plus dans cette direction.

## 2 Objectifs

L'objectif principal de cette thèse est d'explorer l'utilisation de mécanismes situés à la croisée du calcul neuro-inspiré et de l'apprentissage par renforcement dans le cadre de ces architectures dites *neuromorphiques*.

Le cadre général de l'*apprentissage par renforcement* (ou AR) (Sutton and Barto, 1998) propose un fondement théorique pour la prise de décision dans l'incertain. Au sein de l'équipe BISCUIT, c'est la piste principale que nous explorons pour *guider* les processus d'auto-organisation. Les algorithmes classiques sont majoritairement dédiés à des approches discrètes et centralisées peu compatibles avec nos paradigmes de calcul. De même, les mécanismes actuellement en vogue dans l'apprentissage par renforcement profond s'appuient sur des descentes de gradients et ne peuvent pas s'adapter à nos contraintes de décentralisation et de non-supervision.

C'est pourquoi nous voulons explorer plus avant un des mécanismes d'apprentissage issus du connexionisme où décentralisation et codage par population sont plus faciles. Il s'agit de la *Spike-Timing Dependant Plasticity* (ou STDP) (Markram et al., 1997; Bi and Poo, 1998), une règle d'apprentissage non-supervisée que l'on peut toutefois moduler, par exemple en tenant compte d'un signal de renforcement inspiré de l'apprentissage par renforcement.

Plusieurs travaux ont déjà été menés dans cette direction (voir, par exemple, (Florian, 2007; El-Laithy and Bogdan, 2011)), mais ces travaux sont encore peu nombreux. Qui plus est, ils ont été

---

1. Bio-Inspired Situated Cellular and Unconventional Information Technology, <http://biscuit.loria.fr/>

2. [www.loria.fr](http://www.loria.fr)

menés indépendamment des avancées actuelles sur les processeurs neuromorphiques, dont certaines sont très récentes (carte Intel “Loihi”, (Davies et al., 2018)). Or une tendance forte des processeurs neuromorphiques récents est justement l’implantation sur puce de mécanismes STDP configurables, qui permet l’adaptativité des modèles implantés tout en se basant sur des règles d’apprentissage décentralisées et locales. Le but de cette thèse est donc d’étudier cette famille d’algorithmes en se plaçant dans le cadre des paradigmes de calculs étudiés par l’équipe BISCUIT et dans la perspective de l’émergence des circuits neuromorphiques. Il s’agira notamment de :

- mener une étude bibliographique sur STDP et AR ;
- explorer la capacité des modèles existants de STDP modulés par AR à permettre l’apprentissage de nos modèles neuronaux (principalement carte auto-organisatrices et champs neuronaux dynamiques)
- proposer des adaptations de ces algorithmes compatibles avec les contraintes imposées par les circuits neuromorphiques.
- adapter ces algorithmes à des problématiques phares de l’équipe, en particulier le contrôle dynamique décentralisé de l’allocation de ressources de calcul sur circuits neuromorphiques.

### 3 Conditions de travail et compétences souhaitées

Le doctorant sera accueilli au Loria de Nancy. Il ou elle travaillera sous la supervision de Alain Dutech et Bernard Girau. Une collaboration scientifique avec les autres membres de l’équipe est attendue, ainsi que des discussions scientifiques plus générales et des collaborations avec d’autres membres du laboratoire. La durée prévue du doctorat est de trois ans.

Outre des connaissances avancées de niveau master en informatique, des bases solides sur les concepts mathématiques associés (notamment probabilités et équations différentielles) est attendue. Le sujet demande aussi une certaine appétence pour la conception de circuits numériques et l’intelligence artificielle. Enfin, le ou la candidate, titulaire d’un master en informatique ou équivalent, doit être créatif, curieux et autonome. L’équipe fournira un ensemble d’outils de programmation et tout le soutien humain nécessaire pour les aspects techniques, ce qui permettra au doctorant ou à la doctorante de se concentrer sur les questions scientifiques. Être à l’aise avec la conception logicielle est également requis, la production du code se fera sous linux.

### Références

- Bi, G.-q. and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons : Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24) :10464–10472.
- Davies, M., Srinivasa, N., Lin, T., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S., Liao, Y., Lin, C., Lines, A., Liu, R., Mathaikutty, D., McCoy, S., Paul, A., Tse, J., Venkataramanan, G., Weng, Y., Wild, A., Yang, Y., and Wang, H. (2018). Loihi : A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1) :82–99.
- El-Laithy, K. and Bogdan, M. (2011). A Reinforcement Learning Framework for Spiking Networks with Dynamic Synapses. *computational Intelligence and Neuroscience*, 2011.
- Florian, R. V. (2007). Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, 19(6) :1468–1502.
- Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic sps and epsps. *Science*, 275(5297) :213–215.
- Sutton, R. and Barto, A. (1998). *Reinforcement Learning*. Bradford Book, MIT Press, Cambridge, MA.