

Pour une implémentation dans un CPU ou un GPU, une problématique d'entonnoir de Von Neumann (également appelée Von Neumann bottleneck selon sa dénomination anglaise) apparaît du fait que l'implémentation d'un réseau de neurones profond implique d'utiliser à la fois la ou les mémoires et le processeur alors que ces derniers éléments sont
5 séparés spatialement. Il en résulte un engorgement du bus de communication entre la ou les mémoires et le processeur.

Il est donc souhaitable de développer des architectures matérielles dédiées, rapprochant mémoire et calcul, pour réaliser des réseaux de neurones rapides, faible consommation et capables d'apprendre en temps réel.

10 Il est connu de réaliser des réseaux de neurones sur la base d'une technologie de type CMOS. Il est entendu par le sigle « CMOS », Oxyde métallique semi-conducteur complémentaire (acronyme provenant de l'expression anglaise « Complementary Metal-Oxide-Semiconductor »). Le sigle CMOS désigne aussi bien un procédé de fabrication qu'un composant obtenu par un tel procédé de fabrication.

15 Il est également connu un réseau de neurones reposant sur des technologies de type optique.

Plus précisément, trois propositions d'architecture font l'objet d'études spécifiques : les réseaux à neurones CMOS et synapses CMOS, les réseaux à neurones optiques et synapses optiques et les réseaux à neurones CMOS et synapses memristives. Des
20 synapses memristives sont des synapses utilisant des memristors. En électronique, le memristor (ou memristance) est un composant électronique passif. Le nom est un mot-valise formé à partir des deux mots anglais memory et resistor. Un memristor mémorise efficacement l'information car la valeur de sa résistance électrique change, de façon permanente, lorsqu'un courant est appliqué.

25 Toutefois, selon chacune de ces technologies, chaque neurone occupe plusieurs dizaines de micromètres de côté. Pour les technologies CMOS et optique, chaque synapse occupe également plusieurs dizaines de micromètres de côté. Il en résulte que, sur une surface limitée correspondant par exemple à une puce électronique, le nombre de neurones et de synapses qui peuvent être intégrés est limité, ce qui résulte en une diminution des
30 performances du réseau de neurones.

RESUME DE L'INVENTION

Il existe donc un besoin pour un réseau de neurones présentant des performances améliorées, c'est-à-dire un plus grand nombre de neurones et de synapses.

35 A cet effet, la présente description propose un réseau de neurones comportant des chaînes synaptiques, chaque chaîne synaptique comportant des synapses, chaque