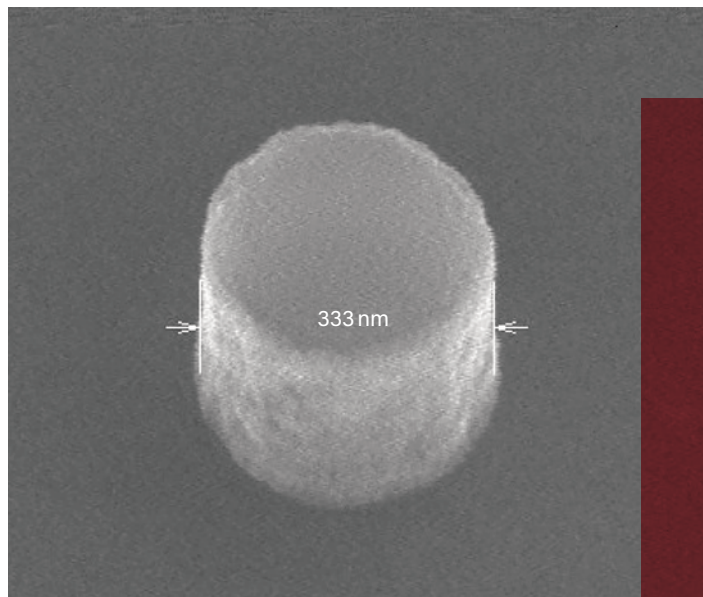




← Les jonctions tunnel magnétiques (*ici vue au microscope électronique à balayage*) sont fabriquées en plusieurs étapes. On fait croître les couches de matériaux l'une après l'autre par pulvérisation cathodique dans une enceinte sous vide. Ensuite, pour obtenir des jonctions d'un diamètre inférieur au micromètre, le matériau est recouvert d'une résine électrosensible et d'un masque sur lequel est dessinée la forme de la jonction. L'ensemble est soumis à une source d'électrons (lithographie) et la résine exposée est conservée. Les parties non protégées par la résine du matériau sont gravées et retirées pour ne garder que la forme de la jonction.

106

sur une puce de silicium. Ces mémoires trouvent déjà des applications en électronique usuelle, car elles associent les qualités des deux grandes familles de mémoires informatiques : la vitesse de lecture et d'écriture des mémoires vives et la capacité des mémoires de stockage à conserver durablement les informations.

On se rapproche de l'idée de puces neuro-morphiques, car il est possible d'insérer facilement des jonctions tunnel magnétiques aux endroits les plus utiles pour stocker des données sous forme de bits au sein des circuits en silicium qui réalisent les calculs. Comme dans le cerveau, où les synapses sont au contact des neurones, ces synapses artificielles seront au plus près du cœur de calcul en silicium. Pour utiliser un algorithme de réseau de neurones dans ce dispositif, il suffit de configurer toutes ces synapses avec les bonnes valeurs pour définir la tâche du programme. Selon le jeu de valeurs mis en place, le réseau sera en mesure d'identifier des éléments dans des images, du son...

En réduisant drastiquement les échanges de données entre mémoire externe et processeur, on exécuterait des algorithmes d'intelligence artificielle en dépensant moins d'énergie. Cette piste est activement explorée dans les laboratoires académiques et industriels. Elle n'est cependant pas entièrement satisfaisante : il faut connaître à

l'avance les bonnes valeurs à mémoriser par les synapses pour effectuer une tâche. Le dispositif n'a pas les capacités d'adaptation d'un vrai cerveau. Or on souhaite, idéalement, un circuit qui soit en mesure d'apprendre de nouvelles tâches.

INITIALISER LES SYNAPSES

Dans les algorithmes de réseaux de neurones, on a souvent une première phase, dite « d'apprentissage », où le réseau ajuste la valeur des synapses afin d'exécuter de mieux en mieux la tâche assignée. Malheureusement, la technique par laquelle l'algorithme identifie les bonnes valeurs des synapses dans les réseaux de neurones artificiels est fondée sur une méthode mathématique complexe, la « rétropropagation du gradient » (*voir les Repères, page 6*). Celle-ci exige des calculs nombreux, très précis et donc énergivores. Elle est bien adaptée sur un ordinateur, mais pas nécessairement pour un système inspiré du cerveau fonctionnant avec des calculs moins précis.

Pour rester dans la logique des systèmes neuro-morphiques, il serait plus judicieux de s'inspirer encore une fois de la façon dont le cerveau apprend. En particulier, les synapses biologiques ne servent pas simplement comme mémoire, elles sont aussi au cœur du mécanisme de l'apprentissage. Elles sont en effet « plastiques » : elles

ajustent leur activité selon la quantité d'information à laquelle elles sont exposées. Par exemple, quand deux neurones connectés par une synapse sont actifs en même temps, cette synapse s'en trouve renforcée avec une activité accrue. Mais ce renforcement n'est pas toujours complet, sinon, d'une certaine façon, chaque nouveau signal effacerait la subtile configuration de la synapse élaborée au cours de toute son expérience passée.

En transposant ce principe aux dispositifs spintroniques, cela reviendrait à modifier la valeur des synapses artificielles en fonction du signal qu'elles reçoivent. Et effectivement, il est possible de reproduire une certaine plasticité avec ces systèmes. Quand une tension électrique est appliquée à une jonction tunnel magnétique, l'effet de transfert de spin met en mouvement l'aimantation d'un de ses nanoaimants. Si la tension est assez élevée et appliquée assez longtemps, ce mouvement entraîne l'inversion de l'aimantation du nanoaimant, et donc de la valeur stockée par la jonction tunnel magnétique. Mais ce renversement n'est pas toujours prédictible : si la tension n'est pas appliquée pendant une durée suffisante, parfois l'aimantation se retourne, parfois non. Plus la tension appliquée à la jonction est élevée, plus la probabilité de basculement de l'aimantation est importante, sans pour autant atteindre 100 %.

UNE CAPACITÉ D'APPRENTISSAGE

Dans une mémoire usuelle, ce comportement incertain doit être évité à tout prix. Pour un système neuromorphique, on obtient des synapses qui, sous l'action des signaux qu'elles reçoivent, transforment progressivement la fonction du réseau de neurones dont elles font partie et adaptent la réponse du système à la nouvelle tâche sans avoir besoin d'utiliser des méthodes aussi lourdes que la rétropropagation du gradient.

En 2015, avec nos collègues, nous avons montré qu'un système équipé de mémoires spintroniques plastiques pouvait apprendre à compter des véhicules sur une autoroute (avec une précision de 95 %) ou à reconnaître des chiffres manuscrits, de la même façon que l'aurait fait un algorithme usuel d'intelligence artificielle. Sur ces tâches relativement simples, les performances s'approchent des techniques plus traditionnelles d'apprentissage des réseaux de neurones (qui reposent cependant sur plus de synapses, avec une consommation plus élevée

en énergie). En revanche, sur des tâches complexes (comme concurrencer un spécialiste du jeu de go), les techniques d'apprentissage bio-inspirées sont pour l'instant loin de rivaliser avec les techniques plus conventionnelles.

Si nous sommes capables de réaliser des synapses artificielles qui stockent de l'information et dont il est possible de moduler la réponse pour simuler la plasticité synaptique et le phénomène d'apprentissage, un système complet neuromorphique doit aussi inclure des neurones pour effectuer les calculs.

Le cerveau humain contient près d'une centaine de milliards de neurones, et les algorithmes d'intelligence artificielle simulent aujourd'hui jusqu'à des dizaines de millions de neurones. Est-il envisageable de réaliser des neurones avec les techniques actuelles de transistors au silicium ? D'abord, si l'on veut autant de composants dans une puce électronique de la taille d'un pouce, il faudrait que chaque neurone ait une largeur bien inférieure à 1 micromètre, ce qui est impossible, car réaliser même un unique neurone artificiel par cette approche nécessite des centaines de transistors.

Par exemple, la puce TrueNorth conçue par IBM en 2014 est une démonstration impressionnante d'électronique neuromorphique qui ne fait pas appel à la spintronique. Ce système comporte cinq milliards de transistors (un nombre considérable), ce qui permet d'implémenter 1 million de neurones et des centaines de millions de synapses. En rapprochant ainsi centre de calcul et mémoire, IBM affirme que la puissance consommée par la puce est inférieure de plusieurs ordres de grandeur à celle d'une puce usuelle.

Avec la spintronique,
il est possible
de reproduire une
certaine plasticité
des synapses,
ici artificielles