Damien Querlioz, University Paris Saclay

ABSTRACT

Memory-Centric Artificial Intelligence with Memory Nanodevices

When performing artificial intelligence tasks, central and graphics processing units consume considerably more energy for moving data between logic and memory units than for doing actual arithmetic. Brains, by contrast, achieve vastly superior energy efficiency by fusing logic and memory entirely, performing a form of "in-memory" computing. Currently emerging memory nanodevices such as (mem)resistive, phase change and magnetic memories give us an opportunity to achieve similar tight integration between logic and memory. In this talk, we will look at neuroscience inspiration to extract lessons on the design of in-memory computing systems. We will first study the reliance of brains on approximate memory strategies, which can be reproduced for artificial intelligence. We will give the example of a hardware binarized neural network relying on resistive memory. Binarized neural networks are a class of deep neural networks discovered in 2016, which can achieve state-of-the-art performance with a highly reduced memory and logic footprint with regards to conventional artificial intelligence approaches. Based on measurements on a hybrid CMOS and resistive Hafnium oxide memory chip exploiting a differential approach, we will see that such systems can exploit the properties of emerging memories without the need of error correcting codes, and achieve extremely high energy efficiency. Second, we will see that brains use the physics of their memory devices in a way that is much richer than only storage. This can inspire radical electronic designs, where memory devices become a core part of computing. We will illustrate this concept by our works using magnetic memories as artificial neurons. We have fabricated neural networks where magnetic memories used as nonlinear oscillators implement neurons, and their electrical couplings implement synapses. We will see that such designs can harness the rich physics and dynamics inherent to magnetic memories, without suffering from their drawbacks. This physics-rich approach nevertheless raises important challenges that we will highlight.

References:

1. M. Bocquet, T. Hirztlin, J.-O. Klein, E. Nowak, E. Vianello, J.-M. Portal and D. Querlioz, "In-Memory and Error-Immune Differential RRAM Implementation of Binarized Deep Neural Networks", IEEE International Electron Devices Meeting (IEDM), p. 20.6.1, 2018.
2. M. Romera, P. Talatchian, S. Tsunegi, F. A. Araujo, V. Cros, P. Bortolotti, J. Trastoy, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. Ernoult, D. Vodenicarevic, T. Hirtzlin, N. Locatelli, D. Querlioz and J. Grollier, "Vowel recognition with four coupled spin-torque nano-oscillators", Nature, Vol. 563, p. 230, 2018.

BIO

Damien Querlioz is a CNRS Researcher at the Centre de Nanosciences et de Nanotechnologies of Université Paris-Sud. He focuses on novel usages of emerging non-volatile memory and other nanodevices, in particular relying on inspirations from biology and machine learning. He received his predoctoral education at Ecole Normale Supérieure, Paris and his PhD from Université Paris-Sud in 2009. Before his appointment at CNRS, he was a Postdoctoral Scholar at Stanford University and at the Commissariat a l'Energie Atomique. Damien Querlioz is the coordinator of the interdisciplinary INTEGNANO research group, with colleagues working on all aspects nanodevice.