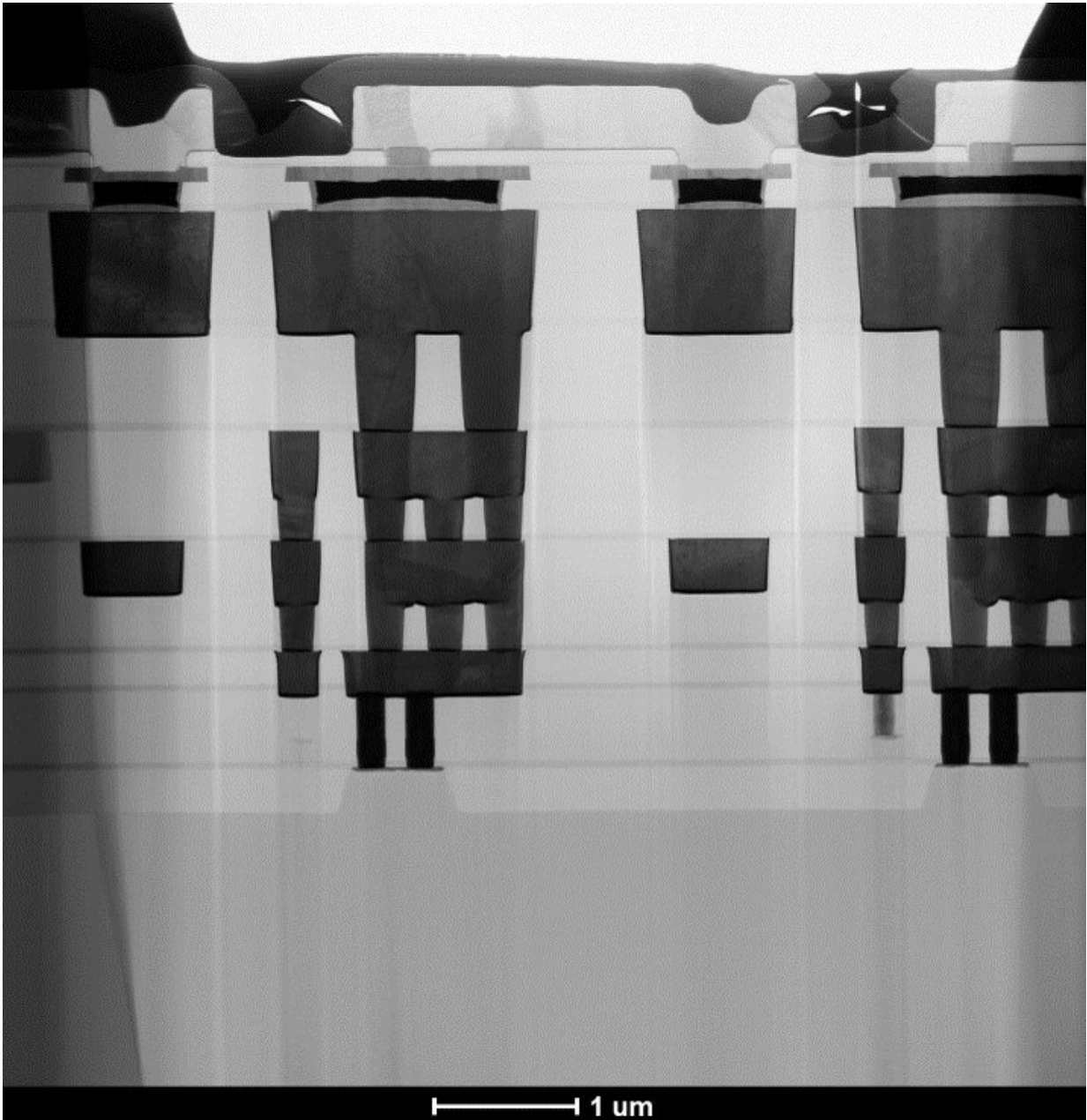# A machine learning framework that exploits memristor variability to implement Markov chain Monte Carlo sampling

February 9 2021, by Ingrid Fadelli

A cross-section of one of the wafers that the researchers used in their experiments. All of the black rectangles in-between the transistor (bottom) and the memristors (top) are four metal layers. In the researchers' chip, these layers were used to connect-up the circuits composed of CMOS transistors and memristors. At the bottom, one can make out the drain and source of a transistor. Credit: Dalgaty et al.

Memristors, or resistive memory devices (RRAM) are nanometer-sized electronic components that can serve as memories for portable computers and other devices. When they are organized in a cross-bar structure, these tiny memories can be used to perform matrix multiplications with low energetic requirements.

Matrix multiplications underpin the operation of all artificial neural network-based models. In fact, when these models operate, input data is multiplied with their "weights matrix."

When a memristor is programmed, its resulting state (i.e., its resistance or conductance) is defined by the random organization of a limited number of atoms inside the device, which link up two electrodes. This means that this state cannot be directly controlled, which significantly limits the potential of memristors for running artificial neural networks.

Over the past few years, engineers have been trying to use series of memristors (i.e., memristor arrays) to allow machine learning models to learn "in-memory." However, the lack of control over their state and the resulting variability in their operation has made this incredibly difficult.

Researchers at CEA-Leti and Université Paris-Saclay-CNRS in France have recently developed a new machine learning scheme that leverages the variability in the state of memristors to implement Markov chain Monte Carlo (MCMC) sampling techniques. These are a class of Bayesian machine-learning algorithms that can derive a sample of a desired distribution of data by constructing what is known as a Markov chain (i.e., a model that represents a sequence of possible events).

"A lot of research and device engineering has gone into making memristors more deterministic during past decades, but they remain pretty random and 'non-ideal'," Thomas Dalgaty, one of the researchers who carried out the study, told TechXplore. "What motivated us (and I

am sure many other research groups working with memristors) was to identify a way to harness this intrinsic device randomness instead of fighting against it. Back in early 2019, we finally realized that a good way to do this might be by turning to the framework of Bayesian machine learning."

Bayesian machine learning approaches make extensive use of "random variables." As a memristor has an intrinsically high variability, Dalgaty and his colleagues realized that it could be considered as a random variable. They thus tried to exploit the implicit "randomness" of memristors to implement an MCMC algorithm.

"Memristors are programmed by 'cycling' devices between a '1'/ON (high conductance) state and a '0'/OFF (low conductance) state," Dalgaty explained. "If we are in the ON state and we switch a device to the OFF state and then back the ON state the conductance of the device measured in the first and second ON states will be different—this is because of the intrinsic random processes that govern the switching physics."

If an RRAM device is switched ON to OFF a significant number of times, by measuring the resulting ON state conductances, this randomness has been noted to follow a predictable "cycle-to-cycle" pattern. More specifically, this "randomness" or variability follows the well-known normal law (i.e., a bell curve), which is applicable to a variety of algorithms, including MCMC sampling techniques.
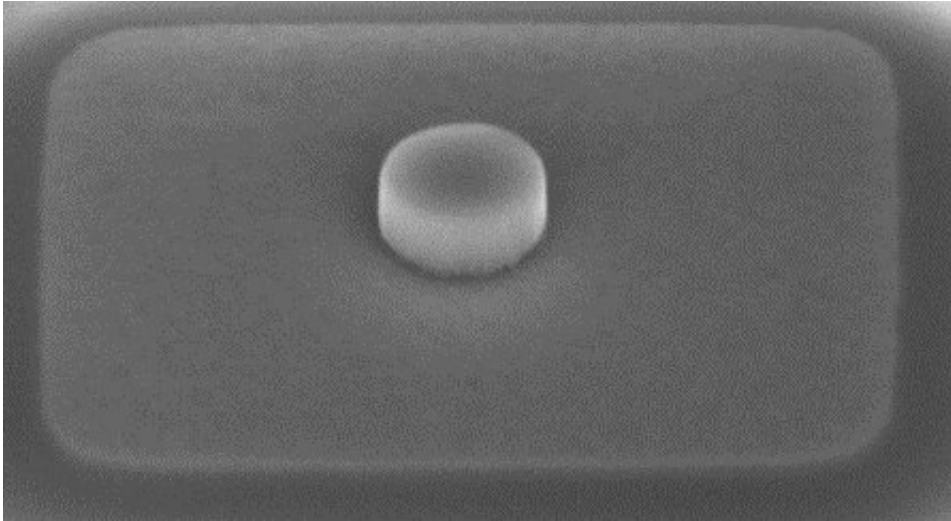
A scanning electron microscopy image of a hafnium-oxide memristor, approximately a hundred nanometres across, after its deposition during the fabrication process. Credit: Dalgaty et al.

"Instead of implementing MCMC on a so-called 'von Neumann' digital computer, which requires lots of energy to move bits between memory and processing circuits on a chip, we can apply voltage pulses over a big array of memristors and let physics do the job for us," Dalgaty said. "The really appealing aspect of this is that we cannot only store information in the memristors, but we can also compute using them, so storage and computing take place in the exact same physical location."

To carry out their experiments, the researchers used memristor chips fabricated at CEA-Leti, a research organization funded by the French government. Remarkably, they successfully utilized these chips to implement a memristor-MCMC algorithm trained to solve different tasks in a laboratory setting.

Dalgaty and his colleagues are among the first to introduce an effective strategy to implement machine learning algorithms in-memory with

memristors. Their approach requires less energy than other techniques based on von Neumann-based digital computers (i.e., a model to design stored-program digital computers).

In fact, implementations based on von Neumann architectures often waste a lot of energy simply to transport bits from a computer's memory to its processor and vice versa. Due to their high energy requirements, when utilizing these strategies machine learning techniques generally need to be ran in the cloud rather than on edge devices.

"We reported a five order of magnitude energy reduction with respect to a von Neumann implementation of MCMC sampling—for context, that is the difference in height between the Burj Khalifa (the tallest building in the world) and a coin," Dalgaty explained. "Because of this massive reduction in the energy required to perform machine learning we think that this 'memristor-MCMC' can help bring machine learning to the edge. Up until now we could not address certain edge applications based on local learning that maybe don't/can't exist yet, due to energy constraints."

To demonstrate the potential of their approach, Dalgaty and his colleagues used memristor-MCMC to solve a task that entailed the detection of heart defects by analyzing electro-cardiogram recordings of patients. Such a technique could ultimately be used to monitor the health of individual patients and prevent or predict the occurrence of heart-attacks, simply by recording and then analyzing data collected by devices worn or implanted close to a patient's heart.

"In our recent paper, we presented our approach using an experimental 'computer-in-the-loop' style setup," Dalgaty said. "This means that we had a computer, connected through a wafer probing station that contacted (with tiny needles) memristor circuits that were designed and fabricated at CEA-Leti. To implement our algorithm, we wrote some

computer programs that generated programming signals that were sent through the needles and onto our circuits."

In their next studies, the researchers will try to devise an integrated circuit (IC) that combines all the functionalities they observed into a single chip, which could be commercialized and used outside of laboratory settings. Overall, designing, fabricating and testing this memristor-MCMC chip might take a few years.

Meanwhile, Dalgaty and his colleagues plan to explore alternative approaches that could enable the implementation of machine learning strategies on edge devices. In their paper, the researchers applied memristor MCMC to a reinforcement learning task.

"Compared to 'supervised' machine learning (that requires a dataset of examples where each example has been labeled with a desired output or class), reinforcement learning doesn't use a labeled dataset and learns directly from interaction with the environment," Dalgaty said. "This reflects much more the reality of local learning the edge, where it will not be possible to have access to pre-labeled datasets. Therefore, another important future direction is to understand how our approach can be best applied in the edge learning setting, investigating not just reinforcement learning as in the paper, but other techniques like 'unsupervised' and 'self-supervised' machine learning too."

More information: In situ learning using intrinsic memristor variability via Markov chain Monte Carlo sampling. *Nature Electronics*(2021). DOI: 10.1038/s41928-020-00523-3.

Citation: A machine learning framework that exploits memristor variability to implement