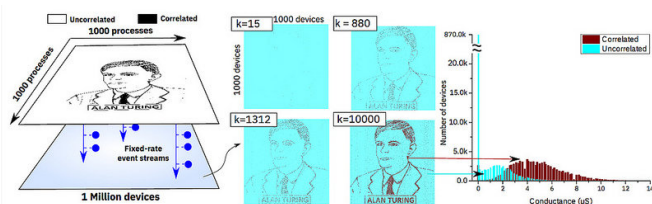


IBM scientists demonstrate in-memory computing with 1 million devices for applications in AI

24 October 2017



A million processes are mapped to the pixels of a 1000 × 1000 pixel black-and-white sketch of Alan Turing. The pixels turn on and off in accordance with the instantaneous binary values of the processes. Credit: Nature Communications

"In-memory computing" or "computational memory" is an emerging concept that uses the physical properties of memory devices for both storing and processing information. This is counter to current von Neumann systems and devices, such as standard desktop computers, laptops and even cellphones, which shuttle data back and forth between memory and the computing unit, thus making them slower and less energy efficient.

Today, IBM Research is announcing that its scientists have demonstrated that an unsupervised machine-learning algorithm, running on one million [phase change memory](#) (PCM) devices, successfully found temporal correlations in unknown data streams. When compared to state-of-the-art classical computers, this prototype technology is expected to yield 200x improvements in both speed and energy efficiency, making it highly suitable for enabling ultra-dense, low-power, and massively-parallel computing systems for applications in AI.

The researchers used PCM devices made from a germanium antimony telluride alloy, which is

stacked and sandwiched between two electrodes. When the scientists apply a tiny electric current to the material, they heat it, which alters its state from amorphous (with a disordered atomic arrangement) to crystalline (with an ordered atomic configuration). The IBM researchers have used the crystallization dynamics to perform computation in place.

"This is an important step forward in our research of the physics of AI, which explores new hardware materials, devices and architectures," says Dr. Evangelos Eleftheriou, an IBM Fellow and co-author of the paper. "As the CMOS scaling laws break down because of technological limits, a radical departure from the processor-memory dichotomy is needed to circumvent the limitations of today's computers. Given the simplicity, high speed and low energy of our in-memory computing approach, it's remarkable that our results are so similar to our benchmark classical approach run on a von Neumann computer."

The details are explained in their paper appearing today in the peer-review journal *Nature Communications*. To demonstrate the technology, the authors chose two time-based examples and compared their results with traditional machine-learning methods such as k-means clustering:

- Simulated Data: one million binary (0 or 1) random processes organized on a 2-D grid based on a 1000 x 1000 pixel, black and white, profile drawing of famed British mathematician Alan Turing. The IBM scientists then made the pixels blink on and off with the same rate, but the black pixels turned on and off in a weakly correlated manner. This means that when a black pixel blinks, there is a slightly higher probability that another black pixel will also blink. The random processes were assigned to a

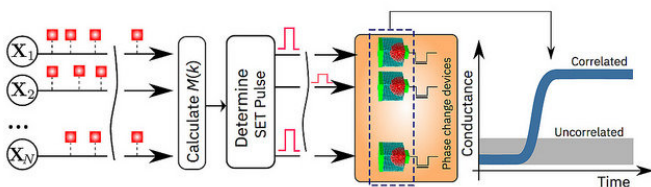
million PCM devices, and a simple learning algorithm was implemented. With each blink, the PCM array learned, and the PCM devices corresponding to the correlated processes went to a high conductance state. In this way, the conductance map of the PCM devices recreates the drawing of Alan Turing. (see image above)

- Real-World Data: actual rainfall data, collected over a period of six months from 270 weather stations across the USA in one hour intervals. If rained within the hour, it was labelled "1" and if it didn't "0". Classical k-means clustering and the in-memory computing approach agreed on the classification of 245 out of the 270 weather stations. In-memory computing classified 12 stations as uncorrelated that had been marked correlated by the k-means clustering approach. Similarly, the in-memory computing approach classified 13 stations as correlated that had been marked uncorrelated by k-means clustering.

More information: Abu Sebastian et al. Temporal correlation detection using computational phase-change memory, *Nature Communications* (2017). DOI: [10.1038/s41467-017-01481-9](https://doi.org/10.1038/s41467-017-01481-9)

Provided by IBM Blog Research

"Memory has so far been viewed as a place where we merely store information. But in this work, we conclusively show how we can exploit the physics of these [memory devices](#) to also perform a rather high-level computational primitive. The result of the computation is also stored in the memory devices, and in this sense the concept is loosely inspired by how the brain computes." said Dr. Abu Sebastian, exploratory [memory](#) and cognitive technologies scientist, IBM Research and lead author of the paper.



A schematic illustration of the in-memory computing algorithm. Credit: IBM Research

APA citation: IBM scientists demonstrate in-memory computing with 1 million devices for applications in AI (2017, October 24) retrieved 21 November 2019 from <https://phys.org/news/2017-10-ibm-scientists-in-memory-million-devices.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.