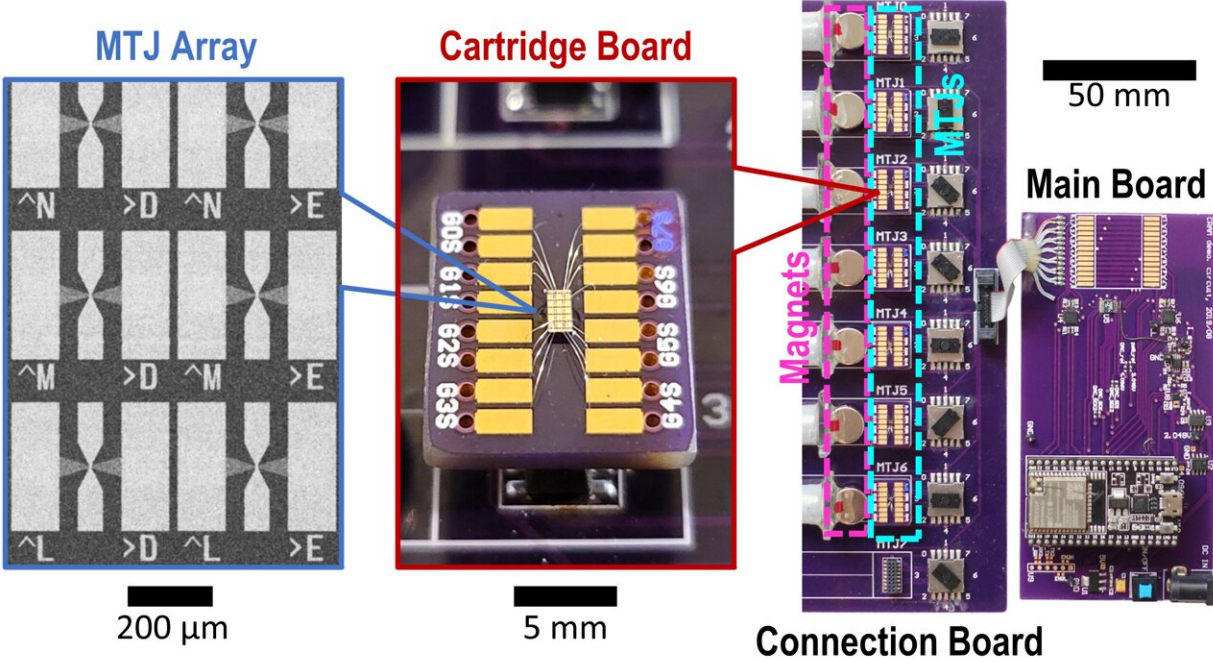


# Engineers develop magnetic tunnel junction–based device to make AI more energy efficient

July 26 2024



A custom-built hardware device plans to help artificial intelligence be more energy efficient. Credit: University of Minnesota Twin Cities.

Engineering researchers at the University of Minnesota Twin Cities have demonstrated a state-of-the-art hardware device that could reduce energy consumption for artificial intelligent (AI) computing applications by a

factor of at least 1,000.

The research is [published](#) in *npj Unconventional Computing* titled "Experimental demonstration of magnetic tunnel junction-based computational random-access memory." The researchers have multiple patents on the technology used in the device.

With the growing demand for AI applications, researchers have been looking at ways to create a more energy efficient process, while keeping performance high and costs low. Commonly, machine or artificial intelligence processes transfer data between both logic (where information is processed within a system) and memory (where the data is stored), consuming a large amount of power and energy.

A team of researchers at the University of Minnesota College of Science and Engineering demonstrated a new model where the data never leaves the memory, called computational random-access memory (CRAM).

"This work is the first experimental demonstration of CRAM, where the data can be processed entirely within the memory array without the need to leave the grid where a computer stores information," said Yang Lv, a University of Minnesota Department of Electrical and Computer Engineering postdoctoral researcher and first author of the paper.

The International Energy Agency (IEA) issued a [global energy use forecast](#) in March of 2024, forecasting that [energy consumption](#) for AI is likely to double from 460 terawatt-hours (TWh) in 2022 to 1,000 TWh in 2026. This is roughly equivalent to the electricity consumption of the entire country of Japan.

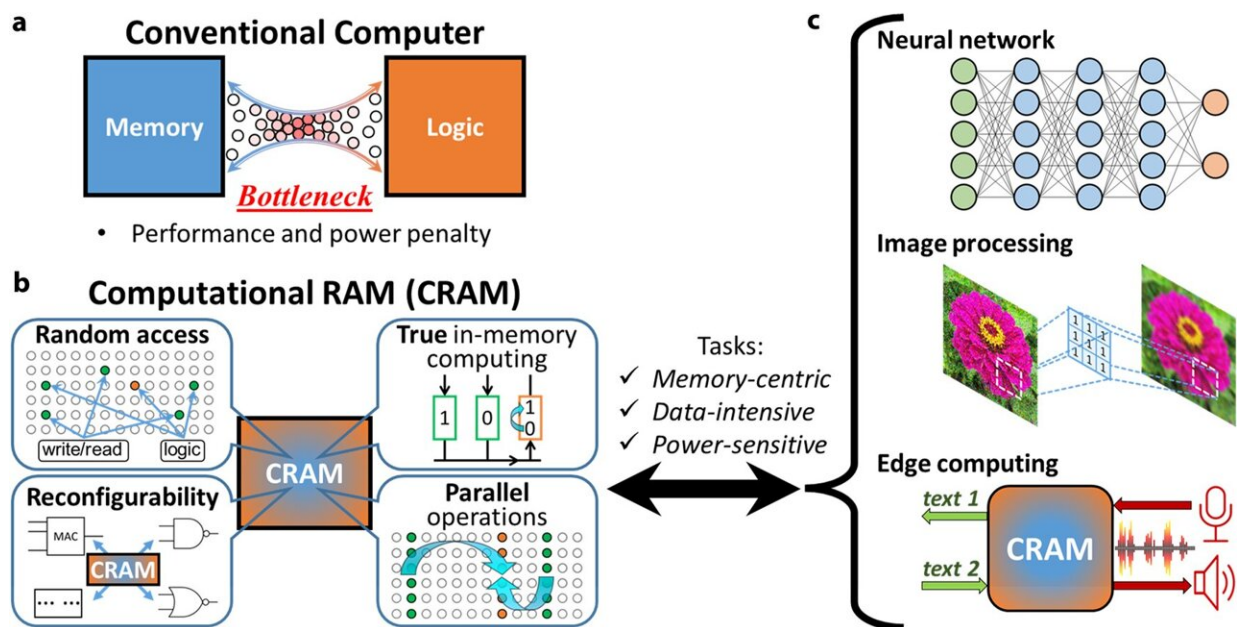
According to the new paper's authors, a CRAM-based machine learning inference accelerator is estimated to achieve an improvement on the order of 1,000. Another example showed an energy savings of 2,500 and

1,700 times compared to traditional methods.

This research has been more than two decades in the making.

"Our initial concept to use memory cells directly for computing 20 years ago was considered crazy" said Jian-Ping Wang, the senior author on the paper and a Distinguished McKnight Professor and Robert F. Hartmann Chair in the Department of Electrical and Computer Engineering at the University of Minnesota.

"With an evolving group of students since 2003 and a true interdisciplinary faculty team built at the University of Minnesota—from physics, [materials science](#) and engineering, computer science and engineering, to modeling and benchmarking, and hardware creation—we were able to obtain positive results and now have demonstrated that this kind of technology is feasible and is ready to be incorporated into technology," Wang said.



Illustrations of CRAM concept, features, and potential applications. Credit: *npj Unconventional Computing* (2024). DOI: 10.1038/s44335-024-00003-3

This research is part of a coherent and long-standing effort building upon Wang's and his collaborators' groundbreaking, patented research into magnetic tunnel junctions (MTJs) devices, which are nanostructured devices used to improve hard drives, sensors, and other microelectronics systems, including Magnetic Random Access Memory (MRAM), which has been used in embedded systems such as microcontrollers and smart watches.

The CRAM architecture enables the true computation in and by memory and breaks down the wall between the computation and memory as the bottleneck in traditional von Neumann architecture, a theoretical design for a stored program computer that serves as the basis for almost all modern computers.

"As an extremely energy-efficient digital based in-memory computing substrate, CRAM is very flexible in that computation can be performed in any location in the memory array. Accordingly, we can reconfigure CRAM to best match the performance needs of a diverse set of AI algorithms," said Ulya Karpuzcu, an expert on computing architecture, co-author on the paper, and Associate Professor in the Department of Electrical and Computer Engineering at the University of Minnesota.

"It is more energy-efficient than traditional building blocks for today's AI systems."

CRAM performs computations directly within [memory cells](#), utilizing the array structure efficiently, which eliminates the need for slow and energy-intensive data transfers, Karpuzcu explained.

The most efficient short-term random access [memory](#), or RAM, device uses four or five transistors to code a one or a zero but one MTJ, a spintronic device, can perform the same function at a fraction of the energy, with higher speed, and is resilient to harsh environments. Spintronic devices leverage the spin of electrons rather than the [electrical charge](#) to store data, providing a more efficient alternative to traditional transistor-based chips.

Currently, the team has been planning to work with semiconductor industry leaders, including those in Minnesota, to provide large scale demonstrations and produce the hardware to advance AI functionality.

In addition to Lv, Wang, and Karpuzcu, the team included University of Minnesota Department of Electrical and Computer Engineering researchers Robert Bloom and Husrev Cilasun; Distinguished McKnight Professor and Robert and Marjorie Henle Chair Sachin Sapatnekar; and former postdoctoral researchers Brandon Zink, Zamshed Chowdhury, and Salonik Resch; along with researchers from Arizona University: Pravin Khanal, Ali Habiboglu, and Professor Weigang Wang.

**More information:** Yang Lv et al, Experimental demonstration of magnetic tunnel junction-based computational random-access memory, *npj Unconventional Computing* (2024). [DOI: 10.1038/s44335-024-00003-3](#)

Provided by University of Minnesota

Citation: Engineers develop magnetic tunnel junction-based device to make AI more energy efficient (2024, July 26) retrieved 27 July 2024 from <https://techxplore.com/news/2024-07-magnetic-tunnel-junctionbased-device-ai.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.