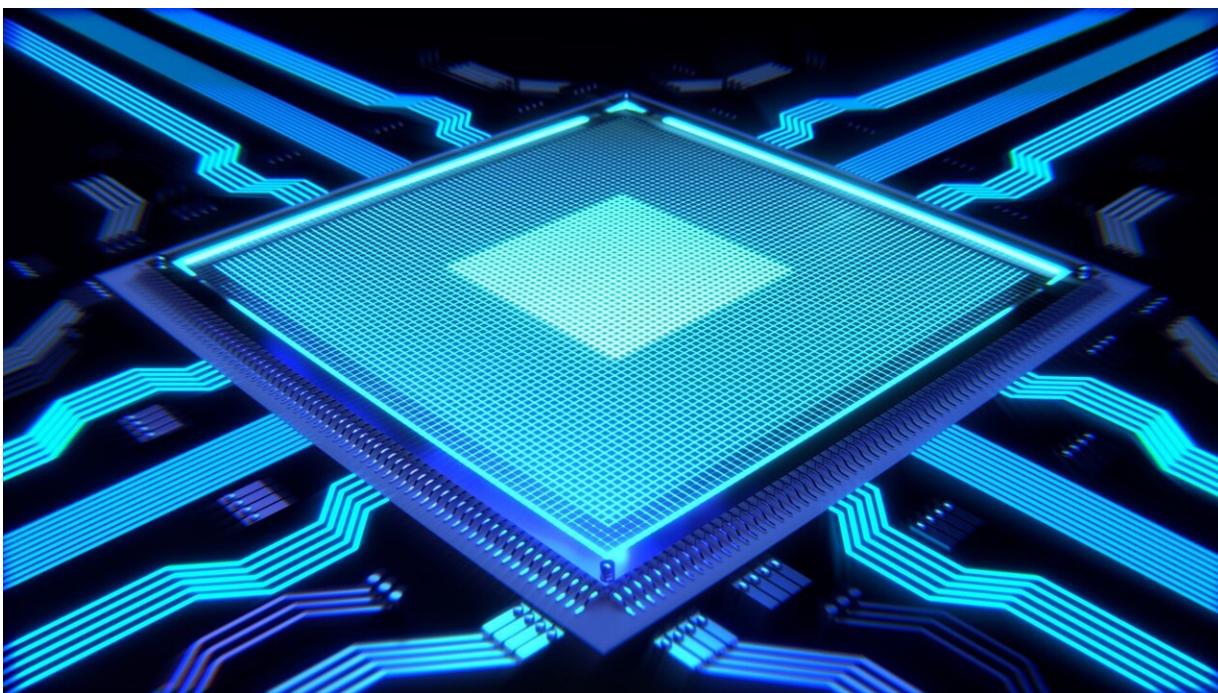


Team creates hybrid chips with processors and memory to run AI on battery-powered devices

January 11 2021



Credit: CC0 Public Domain

Smartwatches and other battery-powered electronics would be even smarter if they could run AI algorithms. But efforts to build AI-capable chips for mobile devices have so far hit a wall—the so-called "memory wall" that separates data processing and memory chips that must work

together to meet the massive and continually growing computational demands imposed by AI.

"Transactions between processors and memory can consume 95 percent of the energy needed to do machine learning and AI, and that severely limits battery life," said computer scientist Subhasish Mitra, senior author of a new study published in *Nature Electronics*.

Now, a team that includes Stanford computer scientist Mary Wootters and electrical engineer H.-S. Philip Wong has designed a system that can run AI tasks faster, and with less energy, by harnessing eight hybrid chips, each with its own data processor built right next to its own memory storage.

This paper builds on the team's prior development of a new memory technology, called RRAM, that stores data even when power is switched off—like flash memory—only faster and more energy efficiently. Their RRAM advance enabled the Stanford researchers to develop an earlier generation of hybrid chips that worked alone. Their latest design incorporates a critical new element: algorithms that meld the eight, separate hybrid chips into one energy-efficient AI-processing engine.

"If we could have built one massive, conventional chip with all the processing and memory needed, we'd have done so, but the amount of data it takes to solve AI problems makes that a dream," Mitra said. "Instead, we trick the hybrids into thinking they're one chip, which is why we call this the Illusion System."

The researchers developed Illusion as part of the Electronics Resurgence Initiative (ERI), a \$1.5 billion program sponsored by the Defense Advanced Research Projects Agency. DARPA, which helped spawn the internet more than 50 years ago, is supporting research investigating workarounds to Moore's Law, which has driven electronic advances by

shrinking transistors. But transistors can't keep shrinking forever.

"To surpass the limits of conventional electronics, we'll need new hardware technologies and new ideas about how to use them," Wootters said.

The Stanford-led team built and tested its prototype with help from collaborators at the French research institute CEA-Leti and at Nanyang Technological University in Singapore. The team's eight-chip system is just the beginning. In simulations, the researchers showed how systems with 64 hybrid chips could run AI applications seven times faster than current processors, using one-seventh as much energy.

Such capabilities could one day enable Illusion Systems to become the brains of augmented and virtual reality glasses that would use [deep neural networks](#) to learn by spotting objects and people in the environment, and provide wearers with contextual information—imagine an AR/VR system to help birdwatchers identify unknown specimens.

Stanford graduate student Robert Radway, who is first author of the *Nature Electronics* study, said the team also developed new algorithms to recompile existing AI programs, written for today's processors, to run on the new multi-[chip](#) systems. Collaborators from Facebook helped the team test AI programs that validated their efforts. Next steps include increasing the processing and [memory](#) capabilities of individual hybrid chips and demonstrating how to mass produce them cheaply.

"The fact that our fabricated prototype is working as we expected suggests we're on the right track," said Wong, who believes Illusion Systems could be ready for marketability within three to five years.

More information: Robert M. Radway et al. Illusion of large on-chip memory by networked computing chips for neural network inference,

Nature Electronics (2021). DOI: 10.1038/s41928-020-00515-3 ,
www.nature.com/articles/s41928-020-00515-3

Provided by Stanford University

Citation: Team creates hybrid chips with processors and memory to run AI on battery-powered devices (2021, January 11) retrieved 26 April 2024 from
<https://techxplore.com/news/2021-01-team-hybrid-chips-processors-memory.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.