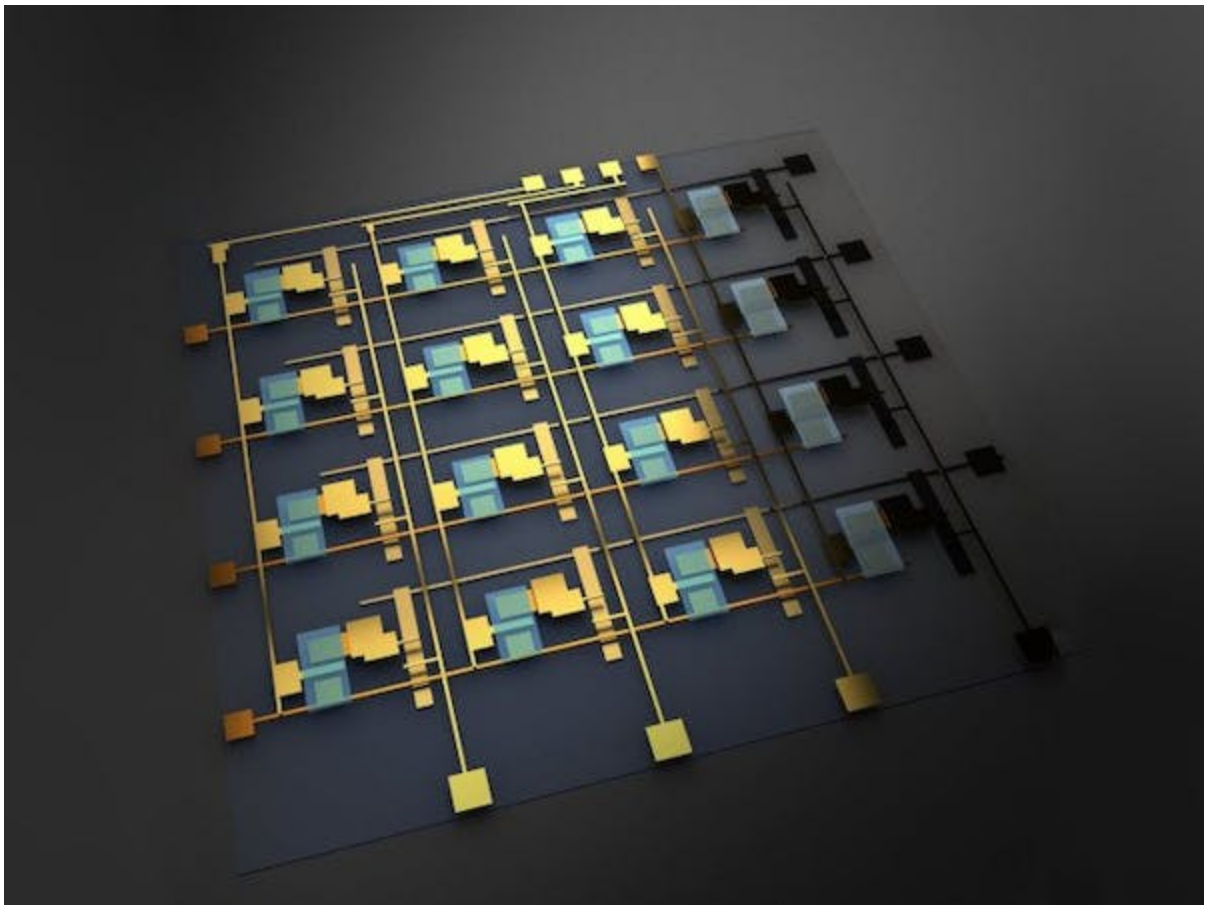
Electrochemical RAM, Bonded on Silicon Chips, Could Boost Edge AI Performance and Efficiency

hackster.io/news/electrochemical-ram-bonded-on-silicon-chips-could-boost-edge-ai-performance-and-efficiency-fda8aa7d4d50

Gareth Halfacree



Researchers from the University of Illinois Urbana-Champaign's Grainger College of Engineering have created what they claim to be the first silicon-integrated electrochemical random-access memory (ECRAM) component ever built — and say it could form the base of a high-efficiency accelerator for machine learning workloads.

"Other ECRAM devices have been made with the many difficult-to-obtain properties needed for deep learning accelerators," admits Qing Cao, professor of the University's Department of Materials Science and Engineering and corresponding author of the work, "but ours is the first to achieve all these properties and be integrated with silicon without compatibility issues. This was the last major barrier to the technology's widespread use."

A new approach to ECRAM production, pictured, allows it to be easily integrated into silicon integrated circuits. (📷: Cui et al)

The team's key breakthrough: being able to integrate ECRAM components, a three-terminal non-volatile memory based on ion exchanges from a reservoir through an electrolyte, with traditional silicon components at the material level — choosing materials which create a functional ECRAM cell yet can be deposited directly onto silicon during chip fabrication.

"While silicon integration is critical, an ideal memory cell must achieve a whole slew of properties," Cao adds. "The materials we selected give rise to many other desirable features," including symmetric injection and removal of ions from the ECRAM's channel to improve reliability and simplify the control scheme, high performance, and longevity beyond 100 million read-write cycles.

ECRAMs are being targeted for use in machine learning accelerator hardware, both storing data and executing calculations directly upon it — an efficiency-boosting approach dubbed compute-in-memory which does away with the bottleneck of having to shuffle data from RAM to the CPU or GPU and back again. As a result, ECRAM-based hardware should offer both improved performance and markedly better efficiency — which has its proponents targeting edge AI implementations once a commercialized version can be produced.

The team has successfully built an 8×8 ECRAM matrix to prove the concept, and is heading to commercialization. (📷: Cui et al)

"Our ECRAM devices will be most useful for AI edge-computing applications sensitive to chip size and energy consumption," Cao explains, pointing to the team's 8×8 ECRAM matrix prototype which can perform matrix-vector multiplication in-memory with high parallelism. "That’s where this type of device has the most significant benefits compared to what is possible with silicon-based accelerators."

The paper detailing the technology, which is being patented by the researchers ahead of work with unnamed "semiconductor industry partners" to bring devices to market, is published in the journal *Nature Electronics* under closed-access terms.