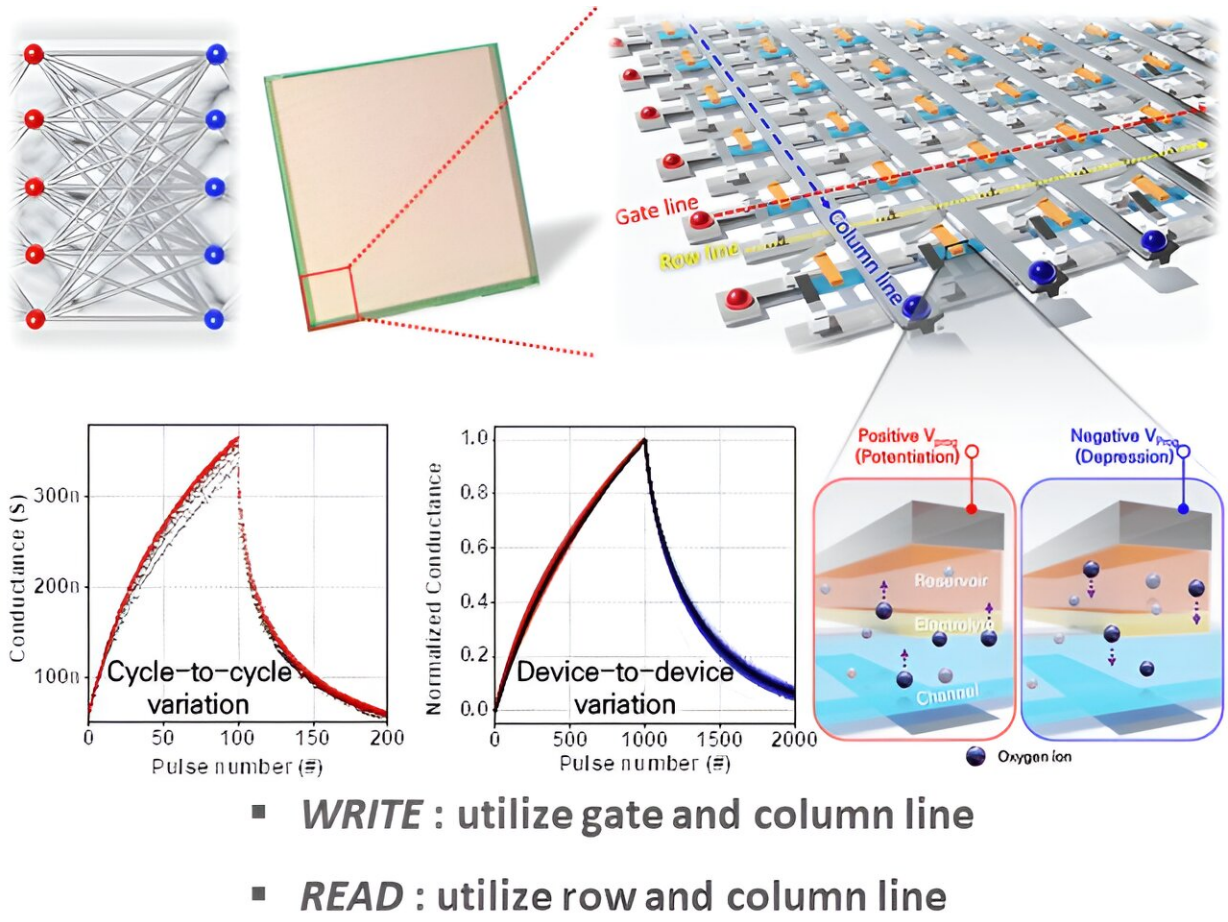
Researchers develop next-gen semiconductor technology for high-efficiency, low-power artificial intelligence

August 1 2024



Cross-point array structure and operation method of the three-terminal ECRAM device fabricated in this study (Top) Measurement results of the array of three-terminal-based electrochemical memory devices, demonstrating excellent characteristics in both cycle and device-to-device scatter, well above the

requirements for training neural networks (Bottom). Credit: POSTECH

A research team has demonstrated that analog hardware using ECRAM devices can maximize the computational performance of artificial intelligence, showcasing its potential for commercialization. Their research has been published in *Science Advances*.

The rapid advancement of AI technology, including applications like generative AI, has pushed the scalability of existing digital hardware (CPUs, GPUs, ASICs, etc.) to its limits. Consequently, there is active research into analog hardware specialized for AI computation.

Analog hardware adjusts the resistance of semiconductors based on external voltage or current and utilizes a cross-point array structure with vertically crossed memory devices to process AI computation in parallel. Although it offers advantages over digital hardware for specific computational tasks and continuous data processing, meeting the diverse requirements for computational learning and inference remains challenging.

To address the limitations of analog hardware memory devices, the research team, consisting of Professor Seyoung Kim from the Department of Materials Science and Engineering and the Department of Semiconductor Engineering and others focused on Electrochemical Random Access Memory (ECRAM), which manages electrical conductivity through ion movement and concentration.

Unlike traditional semiconductor memory, these devices feature a three-terminal structure with separate paths for reading and writing data, allowing for operation at relatively low power.

In their study, the team successfully fabricated ECRAM devices using three-terminal-based semiconductors in a 64×64 array. Experiments revealed that the hardware incorporating the team's devices demonstrated excellent electrical and switching characteristics, along with high yield and uniformity.

Additionally, the team applied the Tiki-Taka algorithm, a cutting-edge analog-based learning algorithm, to this high-yield hardware, successfully maximizing the accuracy of AI neural network training computations.

Notably, the researchers demonstrated the impact of the "weight retention" property of hardware training on learning and confirmed that their technique does not overload artificial neural networks, highlighting the potential for commercializing the technology.

This research is significant because the largest array of ECRAM devices for storing and processing analog signals reported in the literature to date is 10×10. The researchers have now successfully implemented these devices on the largest scale, with varied characteristics for each device.

Professor Seyoung Kim of POSTECH said, "By developing large-scale arrays based on novel memory device technologies and developing analog-specific AI algorithms, we have identified the potential for AI computational performance and energy efficiency that far surpass current digital methods."

More information: Kyungmi Noh et al, Retention-aware zero-shifting technique for Tiki-Taka algorithm-based analog deep learning accelerator, *Science Advances* (2024). DOI: 10.1126/sciadv.adl3350

Provided by Pohang University of Science and Technology

Citation: Researchers develop next-gen semiconductor technology for high-efficiency, low-power artificial intelligence (2024, August 1) retrieved 1 August 2024 from https://techxplore.com/news/2024-08-gen-semiconductor-technology-high-efficiency.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.