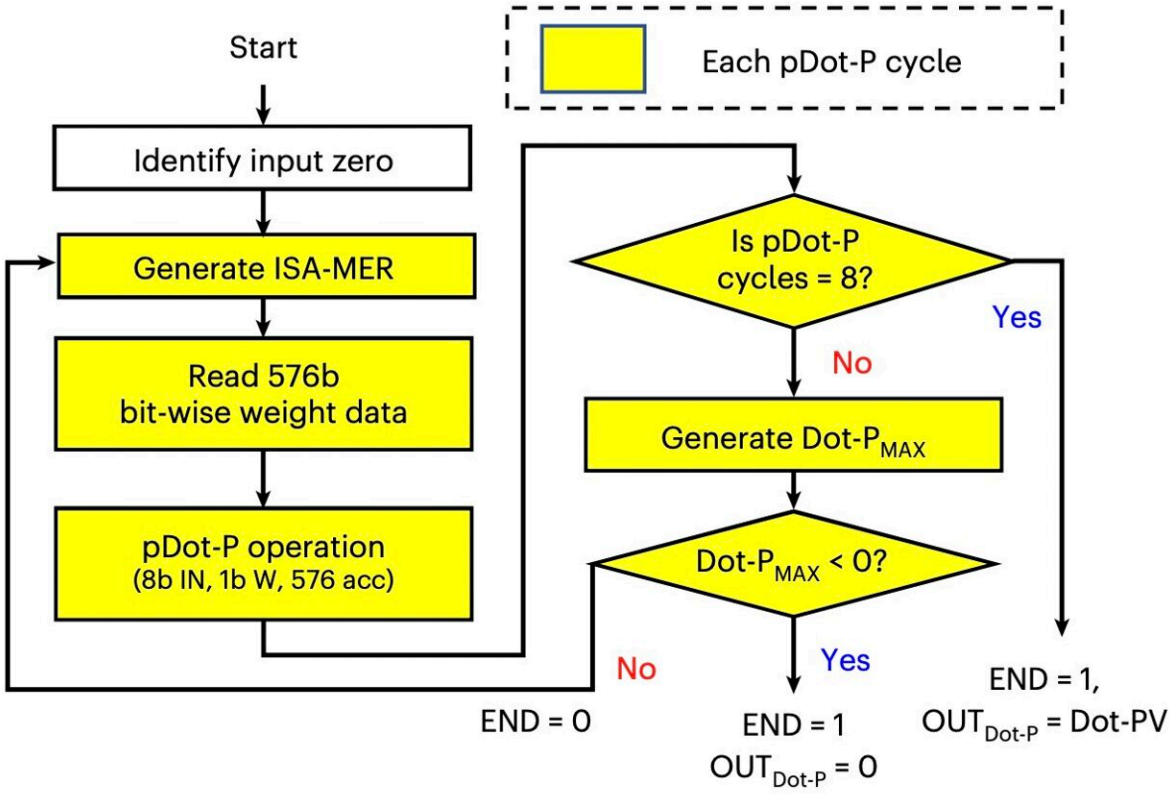
A CMOS-compatible spintronic compute-in-memory macro to secure AI edge devices

July 31 2023, by Ingrid Fadelli



Operation flow chart of the team's nvCIM macro implemented in one Dot-PFC with 8b input–8b weight–26b output. Credit: Chiu et al. (*Nature Electronics*, 2023).

Edge computing applications, which entail the processing and storage of

data at the source of its production (i.e., near where it is created), is now being applied to a growing number of technologies. The application of edge computing translates into devices that can collect, store and process data, such as smart watches, computers that analyze utility grid data, computerized security technologies, and other systems.

As artificial intelligence (AI) algorithms are designed to analyze large amounts of data, they are well-suited for edge computing applications, as they can allow devices to analyze the data they collected and make accurate predictions based on this data. Ideally edge computing devices powered by AI should achieve high prediction accuracy, fast response times and a good power-efficiency, which enables a long battery life.

To facilitate the widespread use of AI-powered edge computing devices and safeguard their users, computer scientists should also ensure that they are protected from cyber-attacks and from the theft of sensitive data. A paper published in *Nature Electronics* introduced a new spintronic compute-in-memory macro that could enhance the security of AI edge devices.

In computer programming, macros are essentially rules, patterns or instructions that outline how input data should be mapped onto a given output. Their macro specifically applies to an on-chip non-volatile compute-in-memory (nvCIM) system, an architecture that combines a processor and a memory component into a single device.

"We report a spintronic nvCIM macro for efficient dot-product edge computing with secure access control for activation, key and data protection against power-on and power-off probing," Yen-Cheng Chiu, Win-San Khwa and their colleagues wrote in their paper.

"The approach relies on spintronic-based physically unclonable functions and two-dimensional half-complement physical encryption, as well as a

snoop-proof self-decryption burst-read scheme in conjunction with a sparsity-and-rectified-linear-unit-aware early-termination compute-in-memory engine."

The nvCIM macro developed by Chiu Khwa and their colleagues can be integrated with existing semiconductor technology, which facilitates its real-world application. The researchers tested its performance in a series of preliminary tests and found that it enabled good protection against malicious attacks, along with rapid response times, and high energy efficiency.

"The 6.6 megabit complementary metal–oxide–semiconductor (CMOS)-integrated macro uses 22 nm spin-transfer torque magnetic random-access memory technology," Chiu, Khwa and their colleagues explained in their paper. "The macro achieves high randomness (inter-Hamming distance, 0.4999) and high reliability for physically unclonable functionality (intra-Hamming distance, 0), as well as a high energy efficiency for dot-product computation (between 30.1 and 68.0 tera-operations per second per watt)."

In the future, the CMOS-integrated spintronic nvCIM macro introduced by this team of researchers could help to improve the security of AI-powered edge computing devices, protecting the sensitive data stored inside them without compromising their speed, accuracy and power-efficiency. In addition, this recent work could inspire other teams worldwide to develop similar solutions to enhance AI edge computing, ultimately contributing to the widespread adoption of AI-supported highly performing technologies.

More information: Yen-Cheng Chiu et al, A CMOS-integrated spintronic compute-in-memory macro for secure AI edge devices, *Nature Electronics* (2023). DOI: 10.1038/s41928-023-00994-0