Energy-Efficient ITO Flash Memory with Superior Endurance for CAM-Based Processing

Jiazheng Chen[1*], Arijit Sarkar[1], Tania Roy[1*]

[1]*Duke University, 101 Science Dr, Durham, NC 27705, US*

3D NAND Flash memory has served as the leading non-volatile storage technology for over a decade for its vertical scalability and excellent data retention [1]. However, its high programming voltage and limited endurance restrict further scaling and adaptation for emerging applications. Ferroelectric FET (FeFET) have emerged as promising alternatives with low write voltages and fast operation, but its narrow memory window and complex fabrication limit large-scale deployment [2]. In this work, we introduce an indium-tin-oxide (ITO)-based Flash memory addresses these limitations (**Fig. 1a**). The proposed device has ZrO_2 as blocking and tunnel dielectric, achieves high endurance exceeding 1,400 DC cycles and 10^5 pulsed cycles, and a low write voltage of 6 V, while remaining compatible with back-end-of-line (BEOL) fabrication processes. Furthermore, we demonstrate its potential for content-addressable memory (CAM) with parallel search functionality.

Multiple devices were measured on the same chip, achieving a yield above 95% with minimal device-to-device variation (**Fig. 1b**). Cycling endurance, evaluated by sweeping the device between –6 V and 9 V for over 1,400 cycles (**Fig. 1c**), demonstrated low cycle-to-cycle variation under continuous DC stress. Pulsing endurance was characterized by repeatedly programming and erasing the memory with ±6 V pulses — less than half of the typical Flash voltage — followed by a low read voltage of 0.5 V after each pulse (**Fig. 1d**). Retention performance was assessed by measuring the device's threshold voltage at various time intervals after a single program/erase pulse (**Fig. 1e**). Finally, we implemented an area-efficient CAM cell using only two Flash memory devices connected in parallel, in contrast to the traditional 16T SRAM-based design. We experimentally validated the operation of a single CAM cell with both write and search functionalities. During operation, the matchline (ML) is precharged, and data is written by applying a program or erase pulse on the search line (SL) and $\overline{SL}$, encoding complementary values across the two devices. Search "0" is carried out by applying 0 V on SL and a search voltage of 1 V on $\overline{SL}$. The match or mismatch state is then determined by observing the ML resistance, where a mismatch causes discharge and indicates a lower ML resistance, as shown in **Fig. 2**. In summary, ITO-channel Flash memory demonstrates high yield, robust endurance, low-voltage operation and CAM functionality, highlighting its promise for energy-efficient in-memory computing.

References

[1] R. Meyer, Y. Fukuzumi and Y. Dong, 2022 International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2022, pp. 26.1.1-26.1.4, doi: 10.1109/IEDM45625.2022.10019570.

[2] H. Xu, J. Yang, T. Kämpfe, C. Zhuo, K. Ni and X. Yin, IEEE Electron Device Letters, vol. 45, no. 1, pp. 112-115, Jan. 2024, doi: 10.1109/LED.2023.3334756

*Corresponding author: email: jiazheng.chen@duke.edu, tania.roy@duke.edu

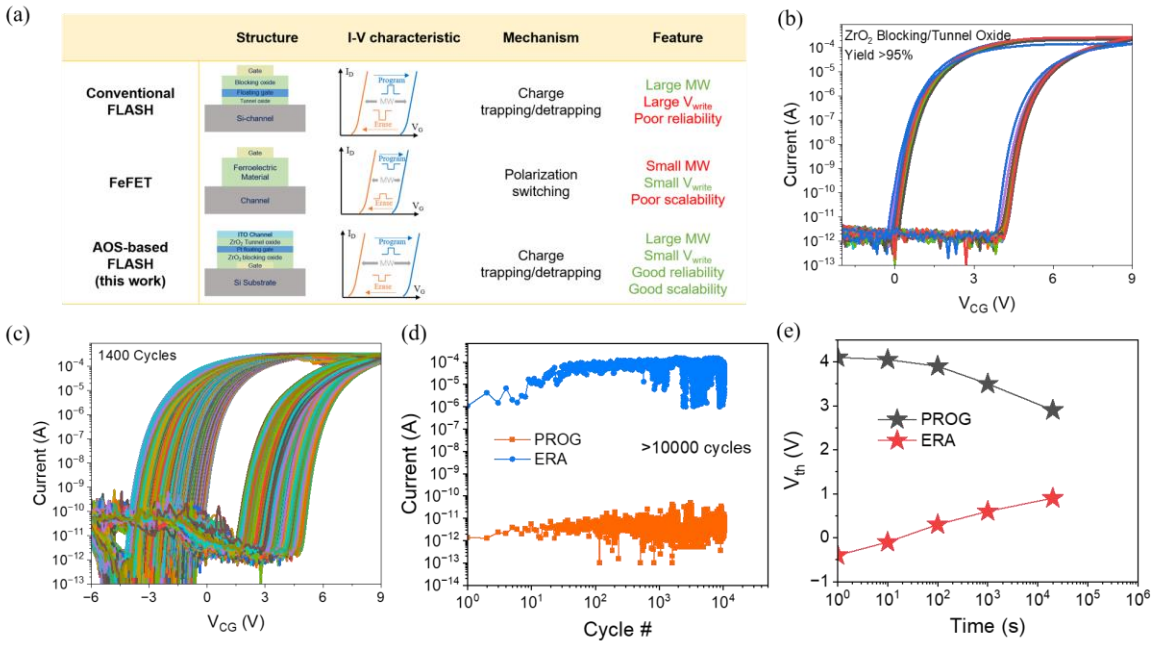

Fig. 1 I–V and memory characteristics of the proposed device. **(a)** Comparison of conventional Flash memory, FeFETs, and the proposed ITO-based Flash memory. **(b)** I–V characteristics across multiple devices, demonstrating a yield exceeding 95%. **(c)** DC cycling endurance over 1,400 cycles. **(d)** Pulsing endurance exceeding 10,000 program/erase cycles. **(e)** Retention performance at room temperature, showing stable threshold voltages in both the erased (low V_{th}) and programmed (high V_{th}) states.

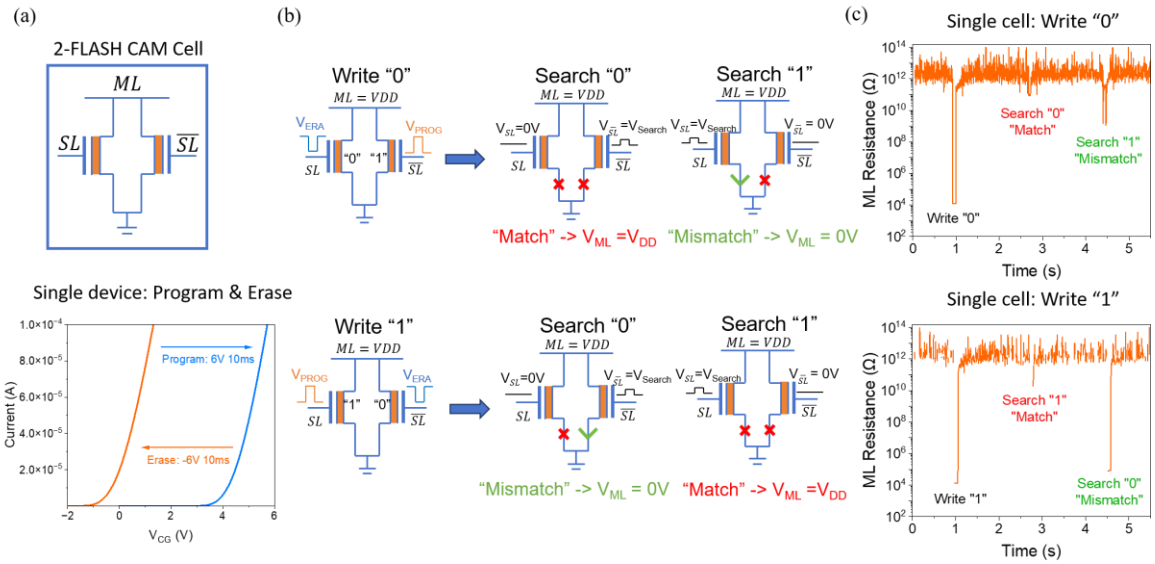

Fig. 2 CAM cell design and experimental demonstration using the proposed ITO-based Flash memory. **(a)** Schematic of the 2-cell CAM architecture with parallel-connected Flash devices and the program and erase I–V characteristics of the Flash devices, with ±6 V pulses. **(b)** CAM operation sequence for write and search operations, match and mismatch states determined by matchline (ML) discharge. **(c)** Measured ML response for "match" and "mismatch" conditions after writing "0" and "1" respectively.