First demonstration of 3D-Stackable Vertical Single-Crystal Si channel FeFET with 10ns Immediate Read-After-Write

Runhao Han[1,2*], Junshuai Chai[1,*], Jia Yang[1], Ruoyao Ji[1], Tao Hu[1,2], Tairan Luo[1,2], Min Liao[1], Xiaoqing Sun[1], Bao Zhang[1], Xin Gong[1], Yong Du[1], Hao Xu[1], Xiaolei Wang[1], Jun Luo[1],Wenwu Wang[1], and Tianchun Ye[1]

[1]Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China.
[2]School of Integrated Circuits, University of Chinese Academy of Sciences, Beijing 100049, China.
*Equal contribution; Email: chaijunshuai@ime.ac.cn, duyong@ime.ac.cn, gongxin@ime.ac.cn

***Abstract*—In this work, we demonstrate a novel dual-layer stacked 3D FeFET with a vertical single-crystal Si channel, fabricated by a CMOS-compatible 3D integration process, with a gate length of 40 nm. The 3D FeFET exhibits an excellent I_{on}/I_{off} ratio of 5.7×10^9 and achieves a memory window of 0.45 V under low-voltage (±3 V) and high-speed (50 ns) operation, with immediate read-after-write delay of 10 ns. The device demonstrates an endurance of 10^5 cycles and no retention loss after 10^3 s at 85 °C. TCAD simulations verify the feasibility of four-layer stacking and reveal sufficient design margin for this architecture. The proposed 3D FeFET offers excellent stackability and scalability, enabling reliable nonvolatile memory and high-speed inference-oriented edge artificial intelligence memory.**

***Keywords*—Ferroelectric, 3D FeFET, Vertical-GAA, high-speed**

I. Introduction

In the era of artificial intelligence (AI), memory devices are required to deliver high density, high speed, and low power consumption, as shown in Fig. 1(a). Ferroelectric field-effect transistors (FeFETs) are promising candidates for inference-oriented edge AI memory due to their non-volatility, low operating voltage, and fast switching speed [1]. However, their practical deployment in large-scale AI systems requires higher integration density. Although conventional planar FeFETs can improve memory density through CMOS-compatible scaling, this approach faces increasing process integration challenges at advanced technology nodes. Therefore, 3D stacking offers a promising solution to further enhance FeFET memory density. Fig. 1(b) shows the schematic of the proposed 3D FeFET.

However, most reported 3D FeFETs are based on oxide semiconductor or polycrystalline silicon channels [2-4]. Oxide channels may suffer from reliability issues. Poly-Si channel FeFETs suffer from high thermal budget requirements, relatively low I_{on}/I_{off} ratio, and large leakage current. Furthermore, high-temperature processing constrains the choice of ferroelectric materials. Thermally robust ferroelectrics, such as HfSiO or HfAlO, require precise control of the doping concentration, which significantly increases process complexity. A single-crystal silicon channel without annealing is an ideal solution to address these problems. However, there is a lack of reports on feasible process integration schemes for 3D FeFETs with single-crystalline silicon channels.

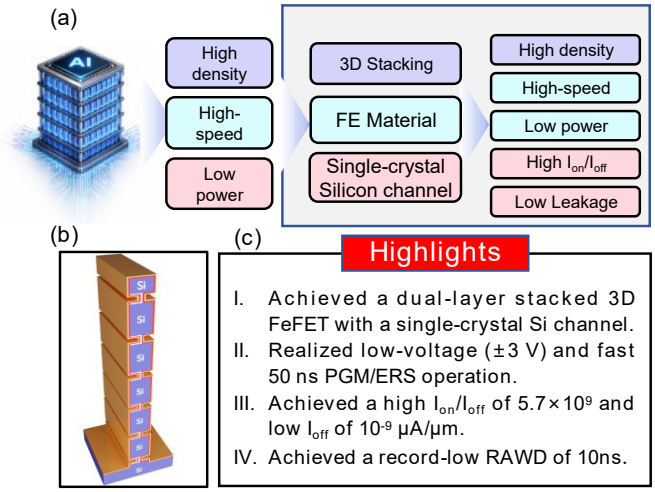

Fig. 1. (a) A 3D FeFET with a single-crystal Si channel can meet the demands of the AI era: high density, high speed, and low power. (b) Schematic of the proposed 3D FeFET. (c) Key highlights of this work.

In this work, we propose a dual-layer stacked 3D FeFET scheme with a single-crystal Si channel enabled by an epitaxial process. This novel 3D FeFET integration approach is fully compatible with Si-based CMOS technology and eliminates the need for high-temperature activation of Si channels. The device achieves a large I_{on}/I_{off} ratio of 5.7×10^9 and a memory window (MW) of 0.45 V under ±3 V, 50 ns. Due to the absence of significant charge trapping, the device achieves an immediate read-after-write delay of 10 ns. It also shows immunity to read disturbance over 10^7 cycles. These characteristics make it well suited for inference-oriented edge AI memory applications. Fig. 1(c) summarizes the highlights of this work.

II. Device and Fabrication

3D FeFETs were fabricated using our previously demonstrated monolithic integration flow [5], as illustrated in Fig. 2(a). Epitaxial Si/SiGe/Si/SiGe/Si stacked films with in-situ doped S/D were grown by reduced-pressure chemical vapor deposition (RPCVD) (Fig. 2(b)). The gate gaps were formed by higher selectivity etching of SiGe with respect to

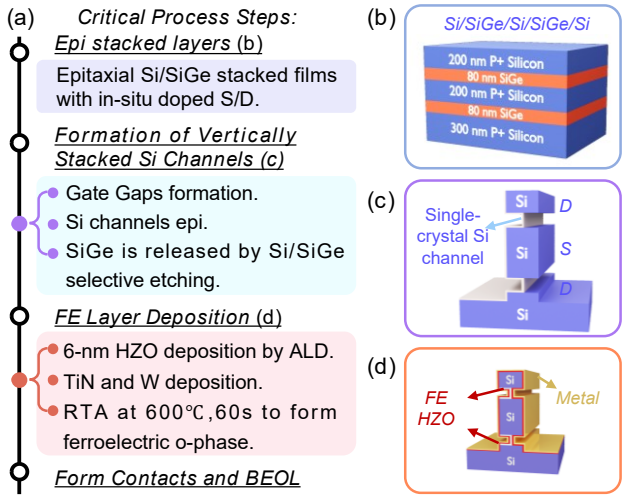

Fig. 2. (a) Key fabrication process for the 3D FeFET. (b) The Si/SiGe stacked films were grown by RPCVD. (c) Forming a single-crystal Si-channel. The channel thickness and gate length were defined by epitaxy. (d) Deposition of the FE stack and metal gate by ALD.

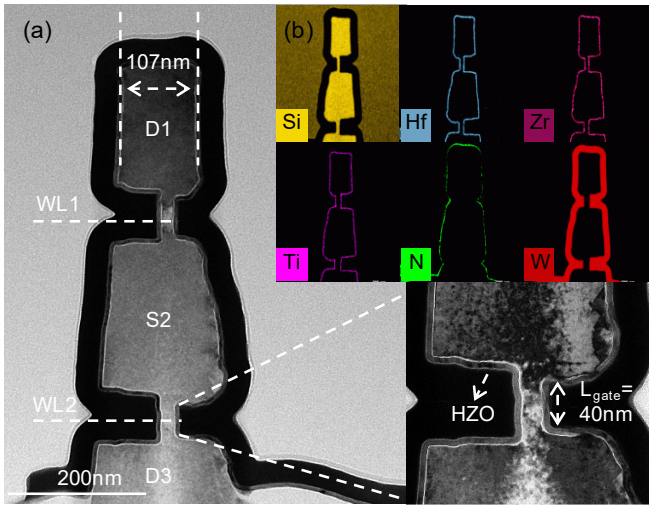

Fig. 3. (a) Cross-sectional TEM image of the 3D FeFET, clearly revealing the dual-layer stacked structure. The gate length of WL2 is 40 nm. (b) Cross-sectional EDS elemental mapping of the 3D FeFET, showing the spatial distribution of the constituent elements.

Si using dry reactive ion etching (RIE). After that, a critical lateral recess step was used to precisely define the recess depth that shapes the epitaxially single-crystal Si channels (Fig. 2(c)). Subsequently, a 6 nm ferroelectric HZO layer was deposited by atomic layer deposition (Fig. 2(d)). The TiN and W were formed as the metal gate. Then, the devices were annealed at 600 °C in N_2 for 60 s to achieve the FE orthorhombic phase. Finally, contacts were fabricated via lithography and anisotropic RIE.

Fig. 3(a) shows the High-Resolution TEM (HRTEM) image for a 2-tier FeFET with L_{gate}=40 nm. WL1 represents the top-tier device, while WL2 represents the bottom-tier device in the vertically stacked structure. Fig. 3(b) shows the EDS elemental distribution of the device. The ferroelectric layer and the gate metal exhibit uniform and conformal coverage. Figs. 4(a)–(c) present the HRTEM images and corresponding FFT patterns of the WL1, WL2, and D3 regions, respectively. Due to the high-quality epitaxial process, the source, drain, and channel regions in both tiers show well-defined single-crystalline characteristics.

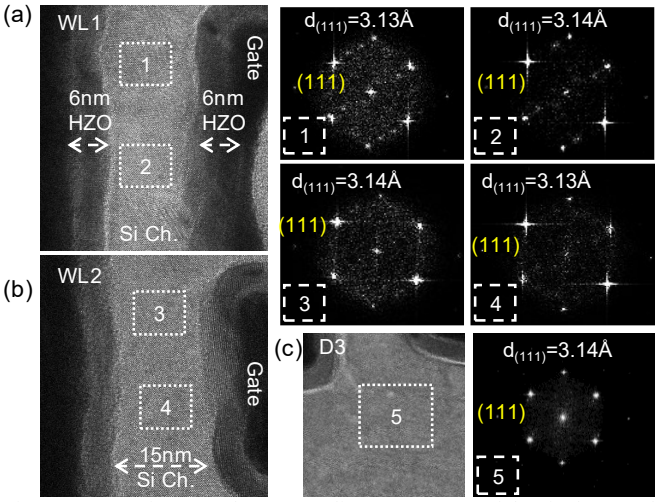

Fig. 4. HRTEM image and corresponding FFT pattern of the single-crystal Si channel in the (a) WL1, (b) WL2, and (c) D3 region. These results confirm that the channels in the proposed 3D FeFET exhibit high-quality single-crystal characteristics.

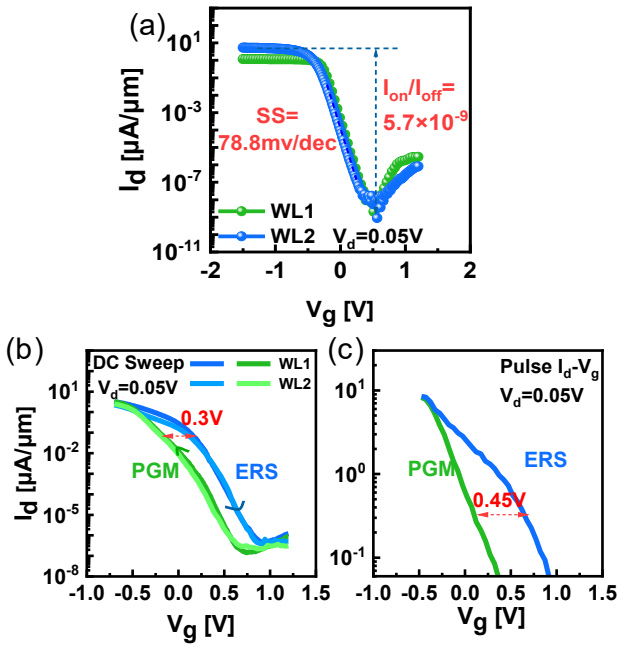

Fig. 5. (a) Id–Vg characteristics of the WL1 and WL2 devices. The devices exhibit an I_{on}/I_{off} of 5.7×10^9 at the initial state. (b) DC-sweeping I_d–V_g curves with P/E voltages of ±3 V and 50 ns pulses. An MW of 0.3 V is obtained, confirming the FE property. (c) Pulsed I_d–V_g curves under the same conditions. The device exhibits an enlarged MW of 0.45 V.

III. Characterization of 3D FeFET

A. I_d-V_g characteristics and MW

Fig. 5(a) shows the I_d-V_g sweep at the initial state for the 3D FeFET. The device shows an excellent I_{on}/I_{off} = 5.7 × 10^9 and low I_{off} of 10^{-9} μA/μm due to the single-crystal Si channel. The devices in the WL1 and WL2 layers show identical I_d-V_g curves, demonstrating excellent device uniformity. As charge trapping at the interface can affect cell read-out, we compared the DC and pulsed I_d-V_g measurements. Fig. 5(b) shows the DC sweep I_d-V_g characteristics. A typical clockwise ferroelectric hysteresis with an MW of 0.3 V is obtained under ± 3 V operating voltage with a 50 ns program/erase (P/E) pulse. The devices in WL1 and WL2 exhibit identical MW. Fig. 5(c) presents the pulsed I_d–V_g results under the same conditions, where an enlarged MW of 0.45 V is achieved. The pulsed I_d-V_g

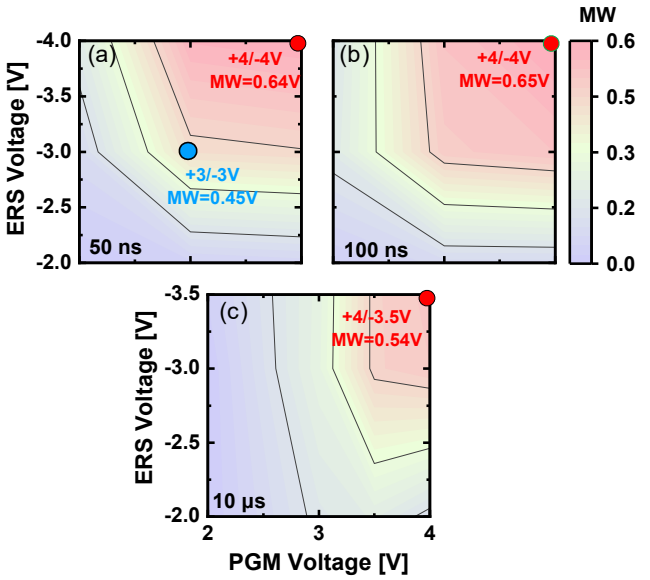

Fig. 6. MW mapping of 3D FeFET devices as a function of the P/E pulse width of (a) 50 ns, (b) 100 ns, and (c) 10 µs.

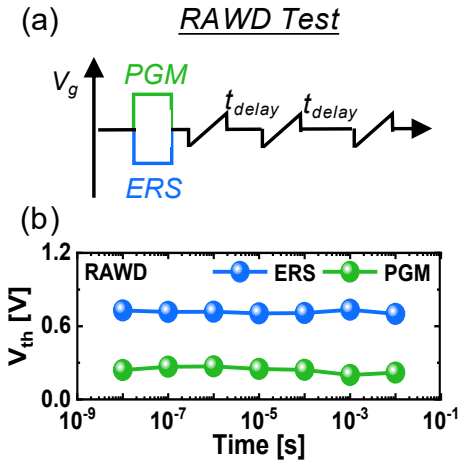

Fig. 7. (a) The test waveform for the read-after-write-delay (RAWD). (b) Measured RAWD for the 3D FeFET. The device exhibits an MW even under a short t_{delay} of 10 ns.

measurement is performed within a µs scale, which suppresses charge trapping and enlarges the MW [2]. The subsequent V_{th} extractions were based on the pulsed I_d-V_g measurements. Fig. 6(a)-(c) summarizes the MW mapping of 3D FeFET devices as a function of the P/E pulse width of 50 ns, 100 ns, and 100 µs, respectively. The device still exhibits an MW of 0.45 V under low voltage of ±3 V and high speed of 50 ns. When the P/E voltage is ±4 V at 50 ns, the MW is enlarged to 0.64 V. With increasing pulse width, the MW shows no further increase. The 50 ns is sufficient to fully switch the ferroelectric polarization.

B. The reliability of 3D FeFET

The immediate read capability and strong immunity to read disturbance are critical for inference-oriented edge AI memory applications [6]. The waveforms of the read-after-write-delay (RAWD) is shown in Figs. 7(a). The P/E voltage is +3/-3 V, 50 ns. Fig. 7(b) demonstrates the RAWD performance of the 3D FeFET, exhibiting a record low read latency of 10 ns. This is essential for fast and real-time computation with 3D FeFETs. The immediate RAWD originates from energetically unfavorable electron trapping in p-FeFETs [7]. Fig. 8(a) presents the read-disturb test waveform. The P/E voltage is +4/-4 V, 50 ns. The disturb

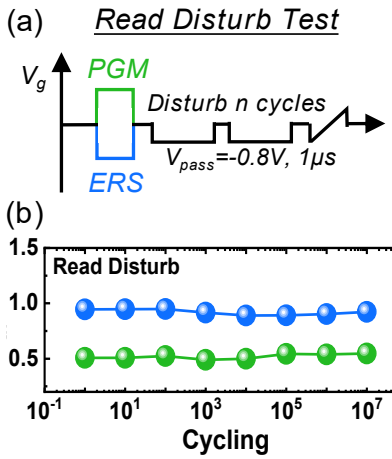

Fig. 8. (a) The read disturb test waveform. (b) Read disturb immunity of the 3D FeFET.

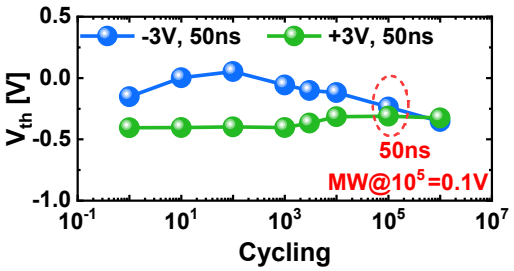

Fig. 9. Endurance measurement of the 3D FeFET under +3/-3 V, 50 ns.

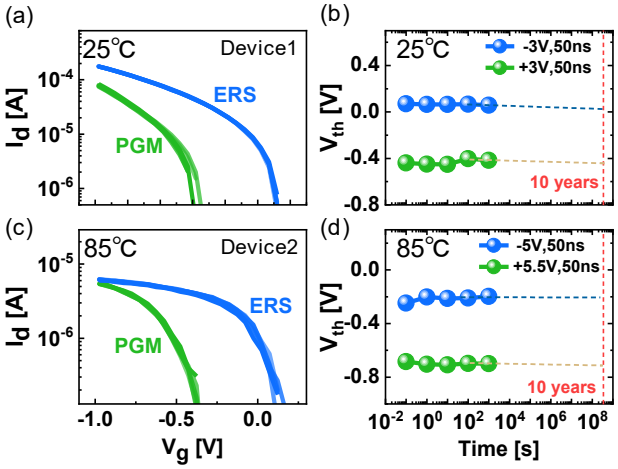

Fig. 10. Retention property of the 3D FeFET measured at 25°C: (a) The I_d–V_g curve and (b) The V_{th} evolution. The linearly extrapolated retention loss (RL) after 10 years is 10%. And retention property measured at 85°C: (c) The I_d–V_g curve and (d) The V_{th} evolution. The device also exhibits good retention performance at 85°C, which can be extrapolated to 10 years.

voltages were set to minimum V_{th} (0.5V) subtract 1.3 V. Fig. 8(b) shows the robust disturbance immunity during 10^7 cycles.

The endurance characteristics are shown in Fig. 9. Under ±3 V operation with a 50 ns pulse width, the device sustains ~1 × 10^5 cycles (Fig. 9). Polarization increases the electric field across the interfacial layer, which limits the endurance of FeFETs to around 10^4~10^5 cycles [8]. The inference-oriented AI workloads have much lower write demand, making 3D FeFETs more suitable for such applications. The retention characteristics at 25°C are shown in Fig. 10(a) and (b), where the device exhibits stable retention with a 10-year extrapolated loss of 10%. Figs. 10(c) and (d) further demonstrate that the retention remains stable at 85°C.

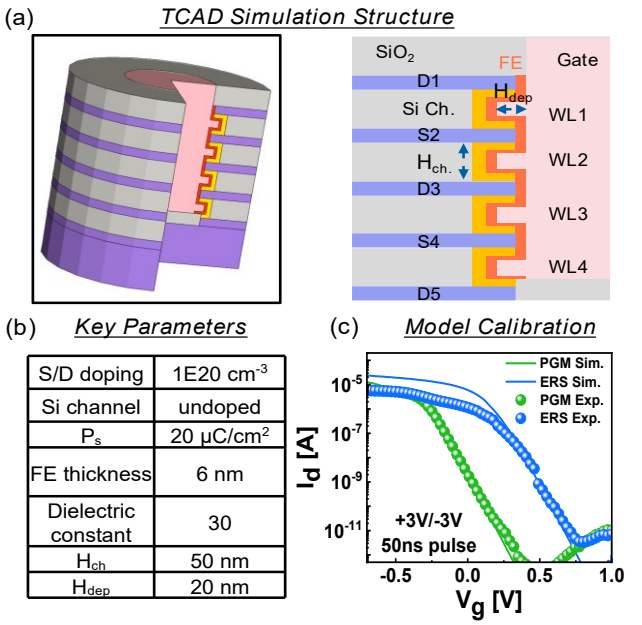

Fig. 11. (a) TCAD-simulated device structure. The horizontal depth (H_{dep}) and height of channel (H_{ch}) are key process parameters, whose impacts on device performance are investigated in the following sections. (b) The key parameters of the simulation in TCAD. (c) The TCAD models are well calibrated with and experimental results.

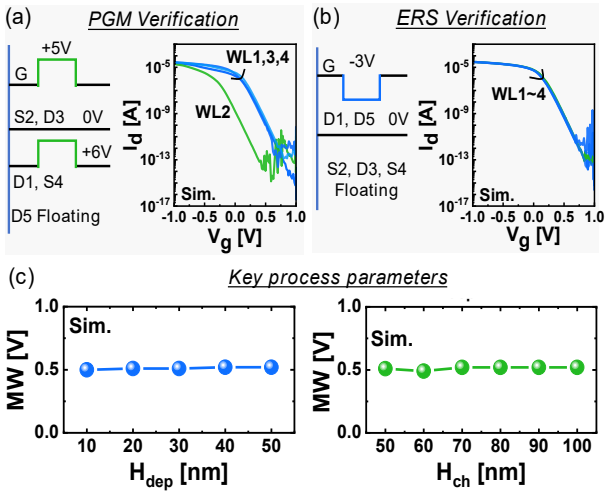

Fig. 12. (a) Verification of the PGM scheme for the 3D FeFET, taking the WL2 cell as an example. The initial state is the ERS state. The applied voltages at each terminal are labeled. The WL2 cell is set to the PGM state, whereas other cells remain ERS state. (b) Verification of the ERS scheme. The initial state is the PGM state. The applied voltages at each terminal are labeled. All cells are switched to the ERS state. (c) Impact of two key process parameters, H_{dep} and H_{ch}, on the MW. The stable MW demonstrates a large design margin for the proposed architecture.

IV. Simulation and Benchmark

Fig. 11(a) shows the TCAD-simulated four-layer stacked 3D FeFET structure. The key device and process parameters are summarized in Fig. 11(b), and the calibrated simulation results are presented in Fig. 11(c). Based on the calibrated TCAD model, the multilayer program and erase behaviors are systematically investigated. Figs. 12(a) and 12(b) illustrate the P/E schemes. Fig. 12(c) shows that variations in the horizontal depth (H_{dep}) and channel height (H_{ch}) have a negligible impact on MW, demonstrating that the proposed architecture provides sufficient design margin for 3D FeFET.

TABLE I. Benchmark of the performance for our 3D FeFET with other 3D FeFET structures.

Only 3D FeFET	2018 IEDM [2]	2025 IEDM [3]	2025 VLSI [4]	This work
3D Stackability	3-tier	2-tier	4-tier	2-tier
Channel Material	Poly-Si	Poly-Si	IGO	Single-crystal Si (First)
$L_{channel}$	50	400	100	40
P/E Voltage	+-10V 100ns	+-2.25V 1us	+4.25/-3.25 50ns	+3/-3V 50ns
I_{on}/I_{off}	10^6	10^5	10^4	5.7×10^9
MW	2V	0.4V	0.5V	0.5V
RAWD	-	-	100ns	10ns
Endurance	10^4	-	10^{11}	10^5
Retention Loss(10 years)	~100%	-	56%	10%

Table I summarizes the benchmark of key performance. The proposed single-crystal-Si-channel 3D FeFET outperforms previously reported 3D FeFETs in I_{on}/I_{off}, write speed, and RAWD.

V. Conclusion

In this work, we experimentally demonstrate a dual-layer stacked 3D FeFET with single-crystal Si channels, an excellent I_{on}/I_{off} ratio of 5.7×10^9, fast P/E speed of 50 ns, and immediate RAWD of 10 ns. This work facilitates the development of high density and speed 3D FeFET for future inference-oriented edge AI memory applications.

Acknowledgment

This work was supported in part by the NSFC under Grant Nos. 92264104 and 52350195.

References

[1] A. Paul, G. Kumar, A. Das, G. Larrieu, and S. De, "Hafnium oxide-based ferroelectric field effect transistors: From materials and reliability to applications in storage-class memory and in-memory computing," *Journal of Applied Physics,* vol. 138, no. 1, 2025.

[2] K. Florent, M. Pesic, A. Subirats, K. Banerjee, S. Lavizzari, A. Arreghini *et al.*, "Vertical Ferroelectric HfO_2 FET based on 3-D NAND Architecture: Towards Dense Low-Power Memory," in *IEDM*, 2018, pp. 2.5.1-2.5.4.

[3] F. Wu, C.-Y. Chiu, C.-H. Wu, Y.-L. Tsai, S.-H. Chen, H.-C. Lin *et al.*, "3D-Stackable Double-Gate Nanosheet (DGNS) FeFETs with Separate Read/Write Paths: Suppressed Variability with Sub-2 % Bit Error Rate and Mitigated Subthreshold Swing Degradation," in *IEDM*, 2025.

[4] Y. Zhou, R. Zhu, W. Luo, X. Xu, S. Qi, Z. Ning *et al.*, "3D NOR-Type FeFETs with Record Endurance of 1011, Fast Erase of 50 ns, and Immediate Read-After-Write for In-Memory Learning," in *VLSI*, 2025, pp. 1-3.

[5] Y. Du, Y. K. Zhang, H. L. Zhu, B. H. Wang, Q. Wang, W. L. Liu *et al.*, "Complementary Vertical FETs (CVFETs) Enabled by a Novel Dual-Side Process," *IEEE EDL,* vol. 46, no. 9, pp. 1473-1476, 2025.

[6] Z. Wang, H. Ma, and J. Zhai, "Low-rank adaptation for edge AI," *Scientific Reports,* vol. 15, no. 1, p. 33109, 2025.

[7] D. Kleimaier, H. Mulaosmanovic, S. Dünkel, S. Beyer, S. Soss, S. Slesazeck *et al.*, "Demonstration of a p-Type Ferroelectric FET With Immediate Read-After-Write Capability," *IEEE EDL,* vol. 42, no. 12, pp. 1774-1777, 2021.

[8] K. Toprasertpong, M. Takenaka, and S. Takagi, "On the strong coupling of polarization and charge trapping in HfO2/Si-based ferroelectric field-effect transistors: overview of device operation and reliability," *Applied Physics A,* vol. 128, no. 12, p. 1114, 2022.