

A 9-Mb HZO-Based Embedded FeRAM With 10^{12} -Cycle Endurance and 5/7-ns Read/Write Using ECC-Assisted Data Refresh and Offset-Canceled Sense Amplifier

Qiqiao Wu^{ID}, Student Member, IEEE, Yue Cao, Qing Luo^{ID}, Senior Member, IEEE, Haijun Jiang, Zhongze Han, Yongkang Han, Chunmeng Dou^{ID}, Member, IEEE, Hangbing Lv^{ID}, Senior Member, IEEE, Qi Liu^{ID}, Member, IEEE, Jianguo Yang^{ID}, Member, IEEE, and Ming Liu^{ID}, Fellow, IEEE

Abstract—Hf_{0.5}Zr_{0.5}O₂ (HZO)-based ferroelectric random access memory (FeRAM) is a good candidate for the embedded nonvolatile memory (eNVM) applications because of its high reliability, high speed, good scalability, and process compatibility with logic large-scale integrated circuits (LSIs). However, challenges still exist in designing robust read/write circuits for high reliability and sufficient read yield. This work presents a 9-Mb (8 + 1-Mb error correcting code (ECC)) HZO-based nonvolatile FeRAM chip with high-performance read and write peripheral circuits. A TiN/HZO/TiN ferroelectric capacitor (FeCAP) is integrated in the back-end-of-line of a 130-nm CMOS process with a 700-nm-diameter capacitor and a mega-level capacity. A temperature-aware ECC-assisted write driver (ECC-WD) is designed to improve the reliability and power efficiency of FeRAM. The offset-canceled sense amplifier (SA) and a dummy-based reference generator are designed to tolerate a small bitline (BL) signal margin and to reduce the read bit-error rate (BER). Measurement results show 2x remnant polarization (P_r) > 30 μ C/cm², >10¹²-cycle endurance, 7-ns write and 5-ns read time, sub-3-V operating voltage, and 10-year data retention at 85 °C.

Index Terms—Ferroelectric capacitor (FeCAP), ferroelectric random access memory (FeRAM), embedded nonvolatile memory (eNVM), refresh, sense amplifier (SA), write driver.

Manuscript received 12 May 2023; revised 26 July 2023 and 17 September 2023; accepted 18 September 2023. Date of publication 11 October 2023; date of current version 29 December 2023. This article was approved by Associate Editor Hidehiro Shiga. This work was supported in part by the NSFC under Grant 92164204, Grant 62222119, and Grant 62025406; and in part by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant XDB44000000. (Qiqiao Wu and Yue Cao contributed equally to this work.) (Corresponding author: Jianguo Yang.)

Qiqiao Wu, Qi Liu, and Ming Liu are with the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100017, China, and also with the Frontier Institute of Chip and System, Fudan University, Shanghai 200437, China (e-mail: liuming@ime.ac.cn).

Yue Cao, Haijun Jiang, and Yongkang Han are with the Zhangjiang Laboratory, Shanghai 201210, China (e-mail: yuecao23@m.fudan.edu.cn).

Qing Luo, Zhongze Han, Chunmeng Dou, Hangbing Lv, and Jianguo Yang are with the Laboratory of Microelectronics Devices and Integrated Technology, Institute of Microelectronics of the Chinese Academy of Sciences, Beijing 100017, China, and also with the School of Microelectronics, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: yangjianguo@ime.ac.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSSC.2023.3320659>.

Digital Object Identifier 10.1109/JSSC.2023.3320659

I. INTRODUCTION

IN RECENT years, big data and artificial intelligence (AI) technologies are developing drastically, leading to growing data-intensive applications, such as deep neural networks or graph analytics. However, the limited memory density of static random access memory (SRAM) restricts the ON-chip memory capacity, failing to keep up with the data storage requirement [1]. In addition, large leakage power consumption becomes a critical issue for maintaining data owing to the volatility of SRAMs [2]. Embedded nonvolatile memory (eNVM) technology can have high memory density and keep data while powering off. Combining with the existing ON-chip SRAM, eNVM is expected to improve overall system performance and power efficiency [3], [4], [5], [6], [7], [8], [9]. NOR flash is currently the mainstream solution for eNVM, and the corresponding fabrication technologies have been mature, but the high power consumption and limited reliability pose an unignorable obstacle to its further development. In addition, the complex cell structure of NOR flash restricts the cost reduction and further scalability [10], [11]. Recently, several new types of memory have emerged, which expands the exploration field in potential eNVM solutions. Magnetoresistive random access memory (MRAM) demonstrates great characteristics in high reliability and mature technology, but its high switching current restricts its usage in low-power designs [12], [13]. RRAM presents an appealing low fabrication cost, but the mechanism of its filament formation inherently causes instability and poor reliability [14]. Phase change random access memory (PCRAM) has moderate reliability and high memory density, but its high power consumption makes it more suitable for large-scale OFF-chip storage, instead of embedded memory [15], [16], [17].

Among the emerging memory technologies, ferroelectric random access memory (FeRAM) is one of the promising candidates for eNVM, because of its high reliability, high speed, and low power. Ferroelectric material has two stable polarization states and is able to achieve polarization switching by external electric field [18]. The first concept of ferroelectric memory implementation goes back to the 1950s,

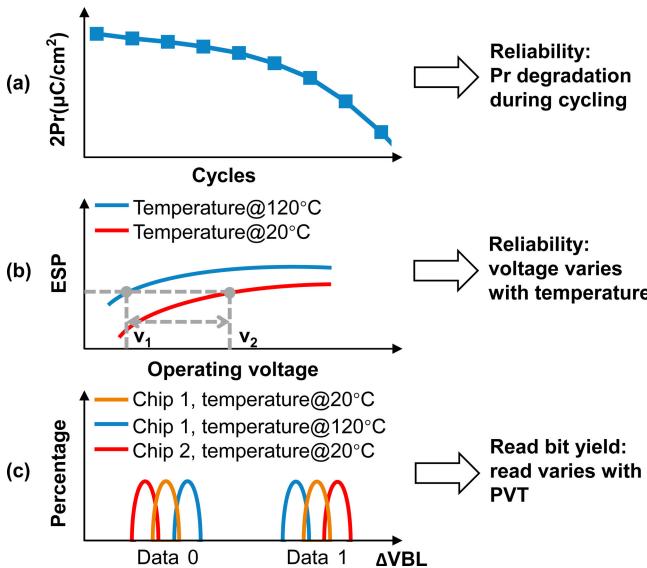


Fig. 1. Challenges in FeRAM design. (a) P_r degradation phenomenon. (b) Effective switching polarization (ESP) changes with temperature. (c) Read errors in FeRAM caused by PVT fluctuation.

and real 2T2C and 1T1C FeRAM macros were reported in the 1980s and 1990s [19]. In the 1990s and early 2000s, FeRAM based on traditional ferro-electric (FE) materials, such as lead zirconium titanate (PZT) or barium titanate (BTO), received high attention for its nonvolatility, low power consumption, high endurance, and compact memory structure like 1T1C dynamic random access memory (DRAM). However, the continuing development has been greatly hindered by the challenges in CMOS integration and scaling to more advanced technology nodes. This situation was changed by the discovery of ferroelectricity in hafnium oxide in 2011 [20]. From that time, the research on $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$ (HZO)-based FeRAM rapidly grows up [21]. In 2019, the first Back end of line (BEOL)-compatible 16-Kb FeRAM array with an HZO-based capacitor was demonstrated [22]. After that, a 64-Kb FeRAM was proposed with an operation voltage of 2.5 V and a read/write speed of 14 ns [7]. These published researches presented the potential of HZO-based FeRAM in embedded memory applications, but reliability and read yield still face fundamental challenges, as shown in Fig. 1.

First, one aspect of the FeRAM reliability problem is caused by the remnant polarization (P_r) degradation phenomenon, which refers to the condition that the $2 \times P_r$ values decrease monotonically with the write cycles. The degradation trend is shown in Fig. 1(a). When the two polarization states become too close and the produced bitline (BL) voltage difference is smaller than the sense margin of its sense amplifier (SA), the stored data could be corrupted. To resolve this issue, recent researchers have focused on improving P_r value. Several articles proposed solutions to optimize the materials for ferroelectric capacitor (FeCAP), including their top/bottom electrodes and the HZO films [23], [24], [25]. Another work demonstrated an FeRAM design with a capacitor-under-bitline (CUB) structure, so that the crystallization annealing temperature can exceed 500 °C without causing metal layer degradation [26]. These proposals provide P_r optimization schemes at device and integration technology scopes, but the circuit design for endurance improvement is still missing.

As presented in Fig. 1(b), the other aspect of the reliability issue comes from the temperature-related P_r variation. Previous research has demonstrated that for $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$ film, the minimum operating voltage for maintaining a certain polarization switching value decreases as temperature increases [27]. The overdrive status may happen when the environment temperature increases while the operating voltage fixes. Such overdrive voltage can lead to additional power consumption and reliability loss.

Second, the read yield is one of the most critical evaluation standards for memory read circuits, as it directly reflects the read quality of a design. However, as demonstrated in Fig. 1(c), the fluctuation of process, voltage, and temperature (PVT) can lead to device characteristic change and mismatch, which may introduce unexpected bias and thus cause read errors. Furthermore, the increasing operating frequency means a shorter time for read window creation, which sets a higher requirement for SA sensing speed. Different studies have presented their solutions to suppress the variation and improve the response time, including offset canceling and precharged sensing [28], [29], [30], [31]. However, the FeRAM-specific research on this issue is at an early stage.

To address the issues above, we demonstrate a 9-Mb HZO-based FeRAM chip with specific read/write peripheral circuits [32]. In this work, a comprehensive analysis of this design is presented, encompassing device integration and peripheral circuit configuration. For the FeCAP device, the integration and measurement details are described. For the write circuit, based on the rejuvenation method proposed in [33], a 2-bit Bose–Chaudhuri–Hocquenghem (BCH) coded error correcting code (ECC)-assisted refresh scheme applies a high refresh voltage to recover P_r when an ECC error is detected. Moreover, a temperature-aware write driver with trimming circuit is designed to generate the required write voltage as the temperature changes, so that power consumption can be reduced by FeCAP overdrive voltage alleviation. For the read circuit, an offset-canceled SA is designed for autozeroing and margin enhancement, which can effectively suppress the negative impact of offset voltage and reduce its standard deviation. In addition, a dummy-based reference generator is employed to compensate for the device changes caused by PVT variations. A 7-ns write and a 5-ns read time, a $>10^{12}$ -cycle endurance, and 10-year data retention at 85 °C are evaluated on the silicon chips. By integrating these high-performance peripheral circuits with well-designed FeRAM array, we announce this first mega-bit-level HZO-based FeRAM chip and verify the feasibility for large-scale embedded FeRAM integration.

The rest of this article is organized as follows: Section II discusses the integration and operation process of 1T1C FeRAM. Section III describes the memory architecture and the high-performance read/write peripheral circuits. Section IV presents the silicon measurement results of the 9-Mb test chip. Conclusions are drawn in Section V.

II. INTEGRATION AND OPERATION OF FeRAM

FeCAP is the key component of an FeRAM cell. Fig. 2(a) shows the mechanism of the HZO-based FeCAP. When the crystallization takes place under mechanical encapsulation, a noncentrosymmetric orthorhombic phase of HZO can be

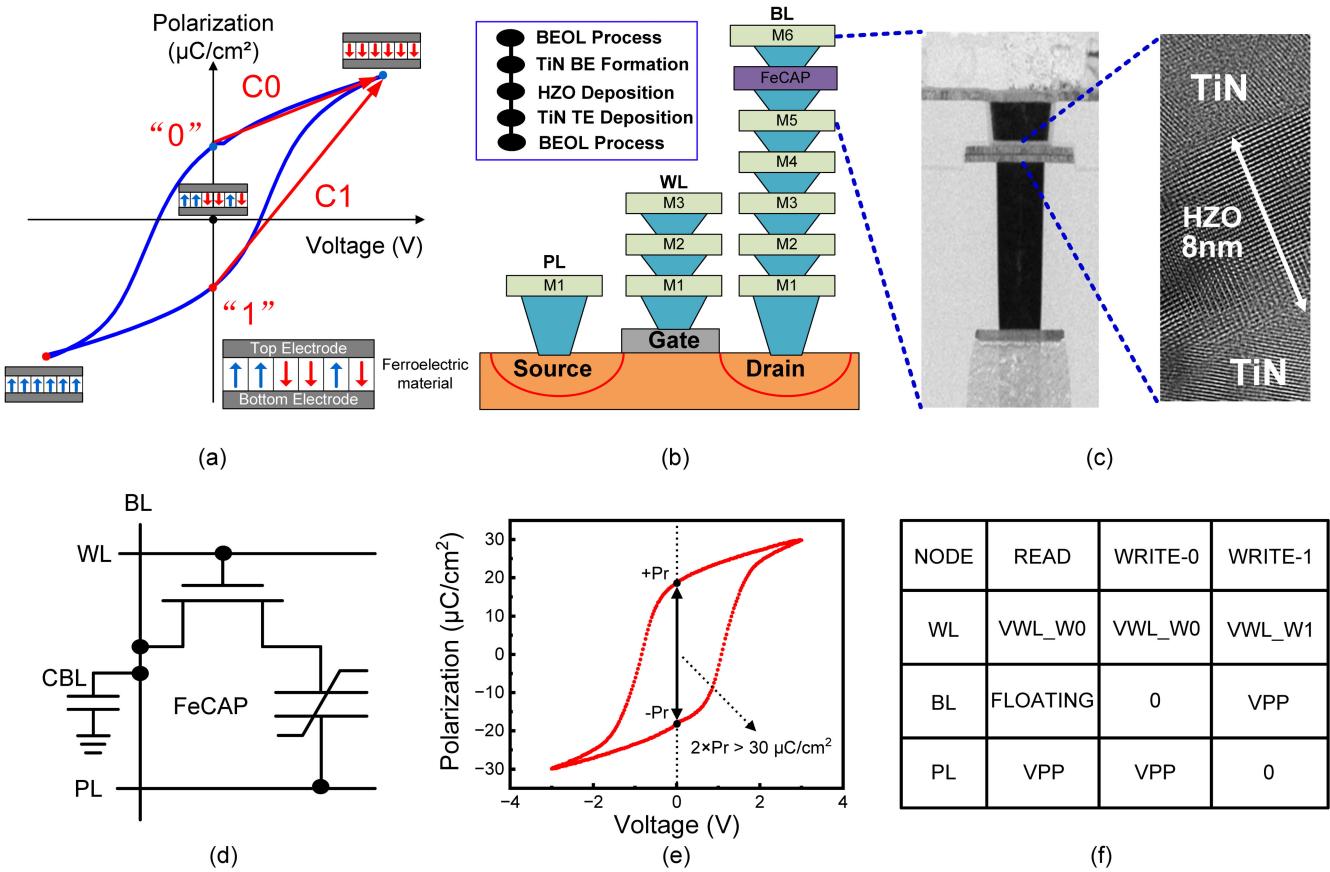


Fig. 2. (a) Mechanism of HZO-based ferroelectric materials. (b) CMOS compatible process flow for 1T1C HZO cell. (c) TiN/HZO/TiN FeCAP layer using 8-nm $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$ and its micrograph. (d) Schematic of 1T1C FeCAP cell. (e) $P-V$ curve measured by FeCAP. (f) Operation table.

formed to enable the electric dipoles in this material [20]. The orientation of such an electric dipole can be switched by an external electric field, so the HZO material will have spontaneous and switchable polarization. As shown in Fig. 2(a), an FeCAP is created using a top electrode, a thin HZO film, and a bottom electrode, and dipoles in the original film are in random directions. By applying a strong positive electric field, such as 2.5 MV/cm, all dipoles can be aligned down and the FeCAP presents a “0” polarization state [7]. Accordingly, dipoles can be switched up by applying a strong negative electric field and the FeCAP presents a “1” polarization state. Such polarization states remain stable after the removal of external field. Here, the field at the polarization switching point is called the coercive field. For reading, a positive field over the coercive field is applied on the FeCAP. If the device is in the “0” state, the polarization change is small, which only induces a low displacement current and releases a small amount of charge (C_0); if the device is in the “1” state, it will be switched to “0” state, and the polarization change will be significant. The large change will cause a high switching current and releases a large amount of charge (C_1). In this work, HZO-based FeCAP is used in large-scale memory and integrated in a silicon chip. Hafnium oxide is a well-established material in the semiconductor industry that is used as a high- k material in CMOS process [18]. HZO-based FeCAP has excellent miniaturization because its thickness can

be reduced to 5 nm [34]. Most importantly, the stable CMOS compatibility of Hf-based FeCAP facilities further researches and potential widespread applications.

Stemming from the ferroelectricity of FeCAP, we constructed the 1T1C nonvolatile FeRAM cell, and its process graph and schematic are presented in Fig. 2(b)–(d). Two additional masks are needed to create an FeCAP in a standard CMOS process. One is to form the TiN/HZO/TiN-stack FeCAP layer, and the other is to pattern the interconnection between the FeCAP electrode and a CMOS via. By restricting the top/bottom electrodes and the HZO film to be the same area and shape, these three layers can share the mask and the fabrication cost can be reduced. For the test chip study, the FeCAP area is designed to be $0.7 \times 0.7 \mu\text{m}$ ($0.49 \mu\text{m}^2$) to reduce design fabrication complexity and improve yield. Meanwhile, the thickness of the HZO layer is selected to be 8 nm to reduce leakage. We have applied the positive-up negative-down (PUND) method for our device measurement, and Fig. 2(e) shows the $P-V$ curve of the HZO-based FeCAP device. Since the capacitors in our design are too small for individual measurement, we designed a test chip to connect the capacitors in parallel and performed the test. A $P-V$ test shows that the $2 \times P_r$ of the FeCAP can reach $>30 \mu\text{C}/\text{cm}^2$, as shown in Fig. 2(e), permitting high-reliability applications.

The operation table of this FeRAM cell is shown in Fig. 2(f), including a bipolar voltage-style writing operation and the

reading operation. For writing 0, the plate line (PL) is connected to a high voltage (VPP), and the BL is connected to ground. When the word line (WL) is active (VWL), the FeCAP is switched to the positive-saturated polarization state, and it returns to the positive-residual polarization state after the WL voltage is removed. For writing 1, the BL is connected to VPP and the PL is connected to ground. When WL is active (VWL), the FeCAP is switched to the negative-saturation polarization state, and it returns to the negative-residual polarization state after the WL voltage is removed. For reading, the BL is left floating, and a pulse is applied to the cell through the PL. The parasitic capacitance of BL (C_{BL}) extracts the ferroelectric charge, via the cell hysteresis, and converts it into a voltage. Since reading the FeCAP state is destructive, the PL needs to be connected to ground. The SA restores the BL to a full-high or full-low state and then realizes the automatic write back of to the FeCAP.

III. CHIP ARCHITECTURE AND CIRCUITS

A. Architecture of the 9-Mb FeRAM Chip

Fig. 3(a) shows the architecture of the 9-Mb FeRAM chip that is comprised of eight 1.125-Mb (excluding reference cells) FeRAM blocks and the peripheral circuits, including the temperature-aware write driver with ECC-assisted refresh scheme, the offset-canceled SA, the dummy-based reference generator, and other necessary modules. As shown in Fig. 3(b) and (c), each FeRAM block has four arrays, each of which is organized as 512 rows by 648 columns with 72 columns for reference generation and 64 columns for ECC parity bit storage. This chip integrates an ECC implementation of a 2-bit error BCH code over 64-bit data with eight parity bits. Each subarray includes 512 rows and 18 columns with two columns for reference generation. Each array is equipped with 32 + 4 data IOs (+4 for ECC) and the two arrays work at the same time to provide 64 + 8 data IOs. Common PLs are shared horizontally by the cells in the same row and vertically within a subarray.

The length of a BL has an important impact on the discharge amount and the voltage window when reading an FeCAP. According to the inspiration from [8], we performed the following simulation test for BL length selection. Fig. 3(d) shows the 3σ voltage signal from the Monte Carlo simulation of the 1T1C cell at 2.5 V. Such simulation applies deviation on the signals to mimic the potential variations in the design and $x\sigma$ is the control parameter of deviation level. By applying a 3σ deviation, we are expected to cover 99.7% of the overall condition. The blue line in Fig. 3(d) is the normalized voltage difference (ΔV) and the red line in Fig. 3(d) is the normalized charge difference (ΔQ) over a range of cell number per BL. For ΔQ , the charge sent by the FeCAP increases with the BL length, since a larger load capacitor can extract more charge during the charge-sharing process. When BL is short and C_{BL} is relatively small, the switching charge released by the FeCAP can quickly raise the BL voltage, leading to the incomplete discharge of the FeCAP for reading “1.” For ΔV , since the incomplete discharge condition above only happens at the state-switching condition, the voltage difference

would increase as C_{BL} increases. When both read conditions enter the full discharge stage, although the FeCAP releases a large amount of charge into a long BL, the large parasitic capacitance of the BL will result in a small voltage difference. The simulation results demonstrate that a good tradeoff in normalized ΔV and normalized ΔQ can be achieved when the number of cells per BL is 512, so the current BL cell number is determined to maintain a balance between the read margin and array efficiency.

B. Write Driver

Fig. 4(a) and (b) illustrates the design motivation of the write driver. According to the measurement results, the minimum required write voltage at -40°C to ensure $2 \times P_r$ over $30 \mu\text{C}/\text{cm}^2$ is 3.5 V. Meanwhile, the required write voltage 120 $^{\circ}\text{C}$ for maintaining the same P_r is 2.0 V, which demonstrates a 43% reduction. If the write voltage is fixed to satisfy the requirement at low temperature, the overdrive write voltage at high environment temperature can cause extra power consumption and potential reliability loss. Fig. 4(b) presents the measurement results of FeRAM refresh effectiveness. When some FeRAM cells experience program and erase cycle failures, they can be repaired by raising the write voltage. In addition, since the raised voltage refresh only happens when a significant degradation is detected and the raise amount is not significant, the occurrence of refresh is limited and its negative impact on strengthening imprint and mechanical stress is minor within the measured cycling period. Therefore, the goal of the write driver is to generate a voltage that tracks temperature and applies a larger write voltage when a refresh is needed.

Fig. 4(c) demonstrates the workflow of ECC-assisted refresh scheme. The ECC scheme in this chip is BCH coding, which can correct multibit errors and is normally used in memory ECC modules [35]. Input data are encoded by the ECC encoder before writing into the array, and if an error is detected by the ECC decoder during data readout, the ECC-based refresh operation will be triggered. By enabling the refresh process on one-bit errors, the occurrence of multibit error, which exceeds the ECC correction capability, can be effectively delayed. To realize the scheme above, the write path of the ECC-assisted FeRAM is depicted in Fig. 4(d). For normal write operation, a 64-bit input data are delivered to the ECC encoder, and a 9-bit trimming data will be generated. The original input data and the trimming data are transferred to the write voltage generator, and by controlling the polarity selector and column selector, the generated write voltage can be applied on the target FeRAM array cells. For refresh operation, the readout data is sent to the ECC decoder. If an error is detected, the decoder will set the error signal to high, triggering refresh write voltage in the voltage generator when the Ref_En signal is high.

Fig. 4(e) shows the schematic of the write voltage generator, which consists of a temperature tracking circuit and a refresh driver circuit. Compared to the bandgap reference (BGR) circuit in [36], we have also included our ECC-assisted refresh driver. While the previous BGR circuit is utilized

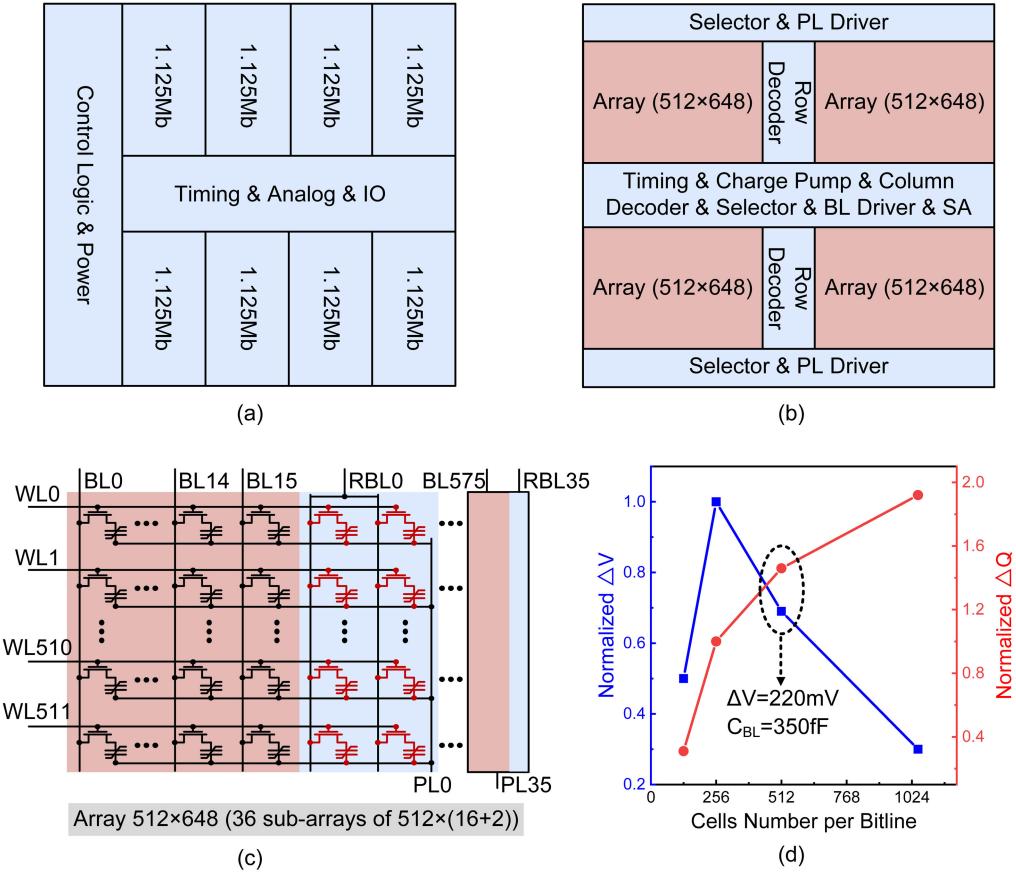


Fig. 3. Architecture of the 9-Mb FeRAM chip, including the architecture of (a) 9-Mb chip, (b) 1.125-Mb block, and (c) 512 × 648 array. (d) Choice of the BL length: 3σ signal from Monte Carlo simulation of the IT1C bitcell at 2.5 V.

to generate read reference voltage, our work integrates the temperature tracker and the refresher to generate the write voltage, so all three trimming resistors are connected to the V_{write} node to simplify overall write voltage control. The temperature tracking circuit consists of two bipolar transistors, $Q1$ and $Q2$, three resistors, $R1-R3$, and an amplifier, $A1$. The base-emitter voltage of a bipolar junction transistor (BJT) exhibits a negative temperature coefficient of about 2 mV/°C, while the thermal voltage (V_T) has a positive coefficient of about +0.085 mV/°C. Under the condition that $R1$ is equal to $R2$, by setting a proper $R2:R3$ ratio, a tracking voltage (V_{track}) can be generated with an expected relation to the temperature as follows:

$$V_{\text{track}} = V_{BE2} + (1 + R2/R3) \cdot V_T \cdot \ln n \quad (1)$$

where V_{BE2} is the base-emitter voltage of $Q2$ and n is the area ratio of $Q2$ to $Q1$. By tuning the value of n , the design in [36] generates the voltage, which is positively related to temperature, while our write driver produces the tracking voltage with a negative temperature coefficient, which enables the implementation of our expected write voltage.

Furthermore, the ECC-based refresh driver circuit includes an amplifier $A2$, three variable resistors, $R4-R6$, and a refresh switch. As shown in this picture, the positive input port of $A2$ connects to the output of the tracking circuit and the negative input port of $A2$ connects to the trimming resistors. By setting $R4-R6$, a normal write voltage is generated when no ECC

errors are detected, and a larger write voltage is produced when ECC failures occur. V_{write} is respectively generated as follows:

$$V_{\text{write}} = (1 + R4/R5) \cdot V_{\text{track}} \quad (2)$$

$$V_{\text{write}} = \left(1 + \frac{R4 \cdot (R5 + R6)}{R5 \cdot R6} \right) \cdot V_{\text{track}}. \quad (3)$$

Fig. 4(e) also shows the 3-bit temperature coefficient trimmer. Each of the resistors $R4-R6$ is programmable with nine resistance values set by three switches, $S0-S2$, for accurately trimming the write voltage, which requires nine trimming bits in total. The Monte Carlo simulation results of V_{write} value versus temperature are presented in Fig. 4(f). Within $\pm 3\sigma$ range, the V_{write} output meets the write voltage requirement in Fig. 4(a).

C. Read Circuits

To enhance the read yield, the offset canceling and dummy-column reference generation schemes are adopted in the read circuit. Fig. 5(a) shows the schematic of the offset-canceled SA, which consists of two cross-coupled inverters (transistors $M1-M4$), four capacitors ($C1-C4$), and eight switches ($SW1-SW8$) for autozeroing and margin enhancement. Furthermore, two dummy cell columns are included to generate dynamic reference voltage for PVT variation tracking.

The read operation is divided into four phases, $P0-P3$, as shown in Fig. 5(b) and (c). During $P0$, the read voltage

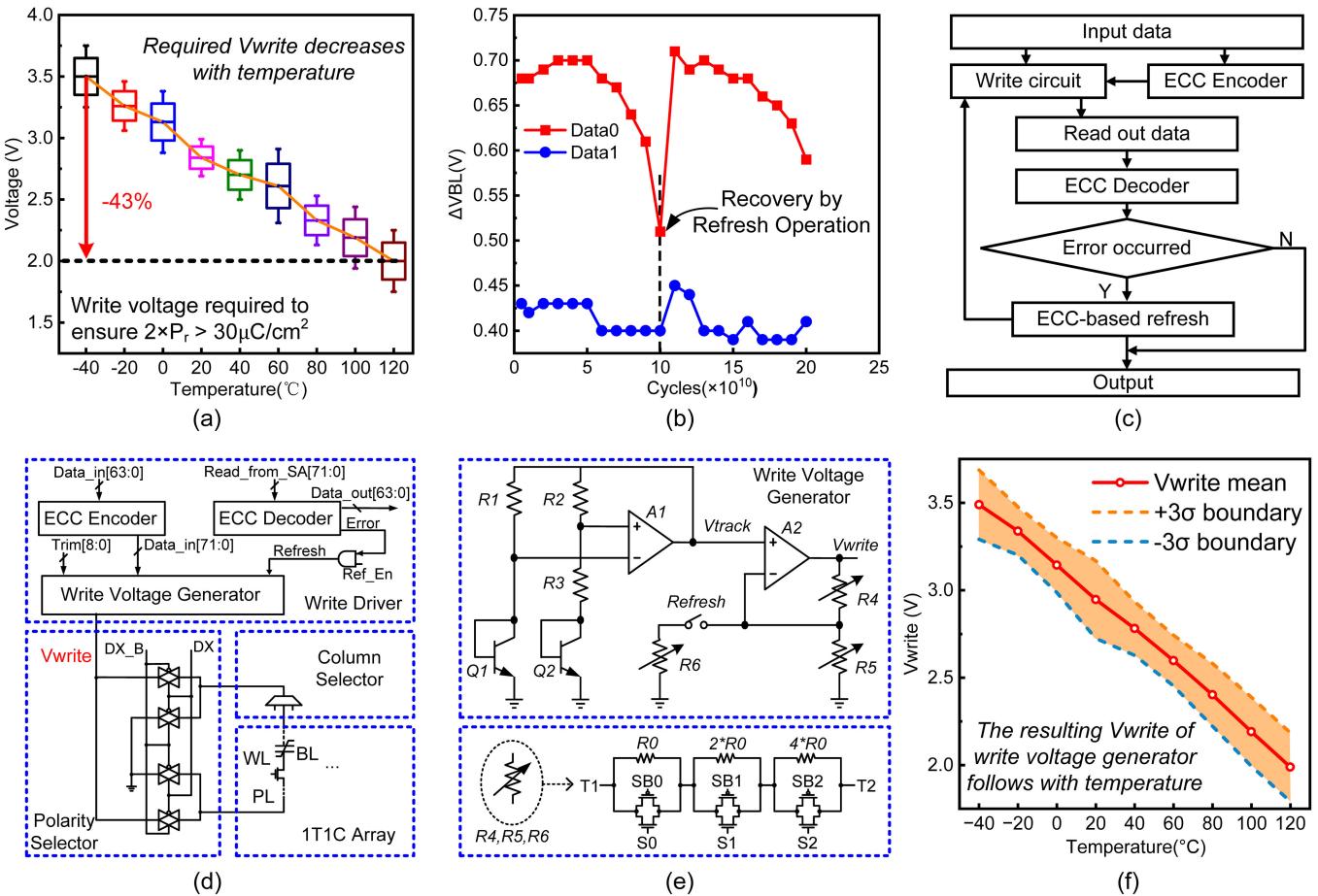


Fig. 4. Design of the write driver. (a) Measurement of required write voltage change with temperature variation. (b) Measurement results of FeRAM refresh effectiveness. (c) Workflow of ECC-assisted refresh scheme. (d) Write path of the ECC-assisted FeRAM, which consists of the write driver, the polarity selector, the column selector, and the 1T1C array. (e) Schematic of the write voltage generator, and the circuit configuration of 9-bit voltage trimmer of R_4-R_6 . (f) Resulting V_{write} generated by the write voltage generator follows with temperature.

V_1 or V_2 is formed on the selected BL, and a reference voltage V_{ref} is formed on the reference BL by two columns of reference cells: one of which always stores a “1” and the other a “0.” The reference voltage tracks process and ferroelectric material variation. In this phase, all switches are open and the inputs of SA are floating. During P_1 , switches SW1, SW4, SW5, and SW6 are on, while the others are off enabling the two cross-coupled inverters to reset to their respective trigger points. During P_2 , switches SW2 and SW3 are on and the others are off, and the difference between the read voltage and the reference voltage ($V_1 - V_{ref}$ or $V_2 - V_{ref}$) is coupled onto Q and QB through capacitors C_1 and C_2 . Then, Q and QB are coupled to QNB and QB. By positive feedback, the voltage difference between Q and QB becomes larger. Finally, the switches SW7 and SW8 are enabled to drive Q and QB rail-to-rail during P_3 .

The measurement of reference voltage with different operation voltages and temperatures is shown in Fig. 6(a) and (b). As shown in these pictures, with the environment temperature of both 25 °C and 105 °C, the proposed reference voltage dynamically tracks BL voltages over a wide voltage range, which maintains a sufficient read window under environment change. Fig. 6(c) shows the Monte Carlo simulations for the proposed offset-canceled SA and the conventional latch-based

SA based on 4000 samples. Simulation results demonstrate that the average offset voltage of the proposed SA is only 18.1 mV, while that of the conventional SA is 44.6 mV. Besides, μ/σ of the proposed SA is 7.2, which is larger than that of the conventional SA. The proposed offset cancellation technique shows good tolerance against the offset and reduces the mean offset voltage by ~60%. Fig. 6(d) shows the comparison of the proposed offset-canceled SA and other state-of-the-art works. Our offset-canceled SA utilizes 12 transistors and four capacitors to achieve an offset standard deviation reduction of 78.86%. Although the extra use of capacitors will lead to an extra area cost, the increase in area is negligible by sharing the same SA through multiple columns of an FeRAM array.

IV. MEASUREMENT RESULTS

Our 9-Mb HZO-based FeRAM chip is fabricated in a 0.13-μm CMOS process, and Fig. 7 shows the micrograph of the proposed chip. The silicon measurement results of the HZO-based 9-Mb FeRAM test chip are described as follows.

Fig. 8(a) and (b) shows the endurance test results, where each dot refers to the results from a single test chip [randomly selected 1-Mb array in (a) and 4-kb array in (b)]. The reason that we do not test the endurance performance of the 9-Mb

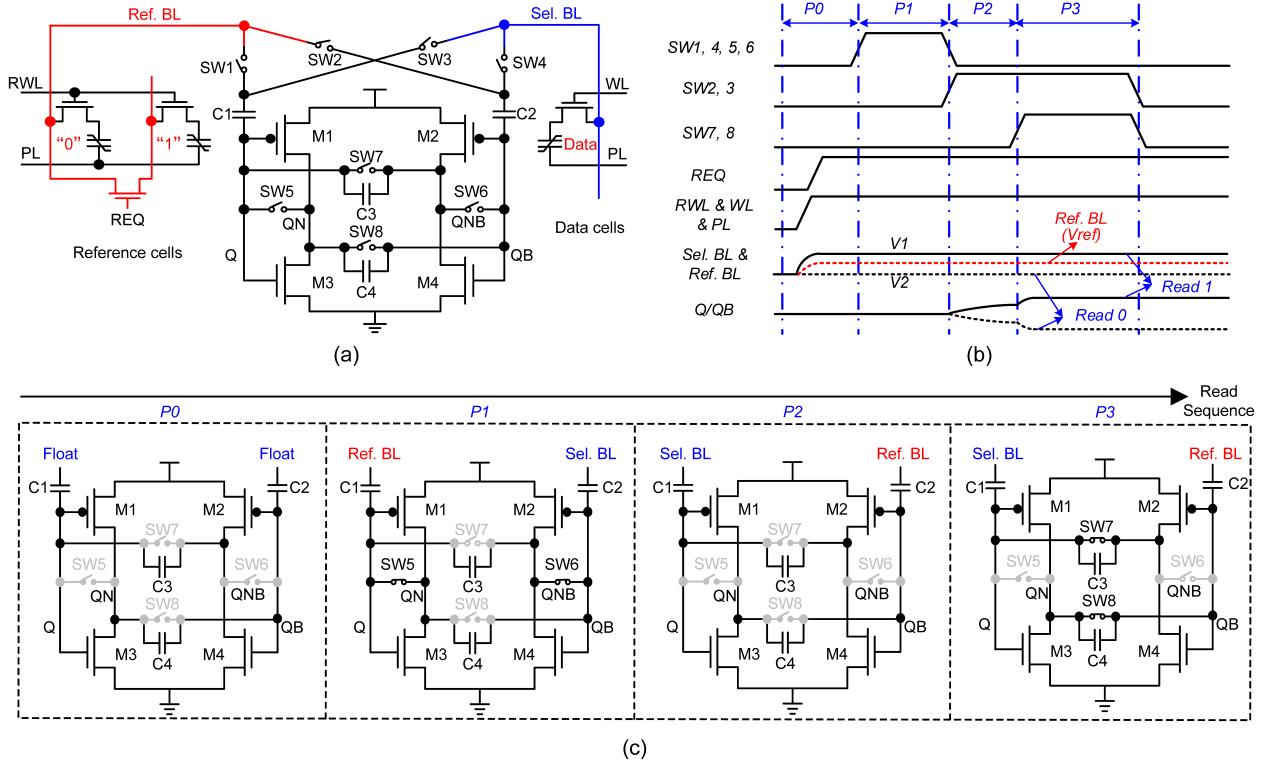


Fig. 5. (a) Proposed offset-canceled SA with the reference and data cells. (b) Work flow during reading. (c) Dynamic states of SA according to read sequence.

TABLE I
COMPARISON TO PRIOR STATE OF THE ART; PRESENTED WORK IS BASED ON CHIP-LEVEL DATA

	This work	ISSCC'21 [3]	ISSCC'20 [4]	ISSCC'17 [5]	IEDM'21 [6]	VLSI'20 [7]	ISSCC'11 [8]	ISSCC'18 [9]	TVLSI'09 [39]	JSSC'10 [40]
Technology (nm)	130	14	22	90	130	130	130	65	130	130
Memory Type	FeRAM (1T1C)	RRAM (1T1R)	MRAM (1T1M)	NOR flash (ESF3)	FeRAM (1T1C)	FeRAM (1T1C)	FeRAM (1T1C)	OSFET (3T1C)	FeRAM (chain)	FeRAM (chain)
Capacity	9Mb	1Mb	32Mb	1Mb	16Kb	64Kb	1Mb	1Kb	64Mb	128Mb
Write Time (ns)	7	-	-	10 μ s	4	14	-	20	-	30
Read Time (ns)	5	5.6	6	11	-	8	200	45	7	30
Retention	10 years /85°C	10 years /85°C	10 years /125°C	10 years	104 s /125°C	100 min /85°C	-	-	-	10 years /85°C
Write Endurance	10^{12}	10^6	10^6	10^4	10^7	10^9	-	10^{14} a	$>10^{10}$	10^{13}
CMOS Compatible	Yes	Yes	Yes	Yes	Yes	Yes	No b	Yes	No b	No b
Scalable	Easy	Easy	Easy	Hard	Easy	Easy	Hard	Easy	Hard	Hard

a. Measured on single cell

b. Pb(Zr_{0.25}Ti_{0.75})O₃ ferroelectric layer used is incompatible with CMOS

chip is that a thorough test on such a chip requires too many write cycles, which will take an unacceptably long test period. As an alternative, we randomly choose multiple chips from different locations in the different silicon wafers for a better coverage rate. To further cover the potential corner cases, we have performed extensive read/write tests on our 9-Mb

chip, and all of the row addresses that we chose to perform the tests are also randomly generated. The bit-error rate (BER) of 12 chips is measured after 10^6 and 10^{12} cycles of writing with and without the ECC-assisted write driver (ECC-WD); no endurance-induced errors were detected using the ECC-WD technique.

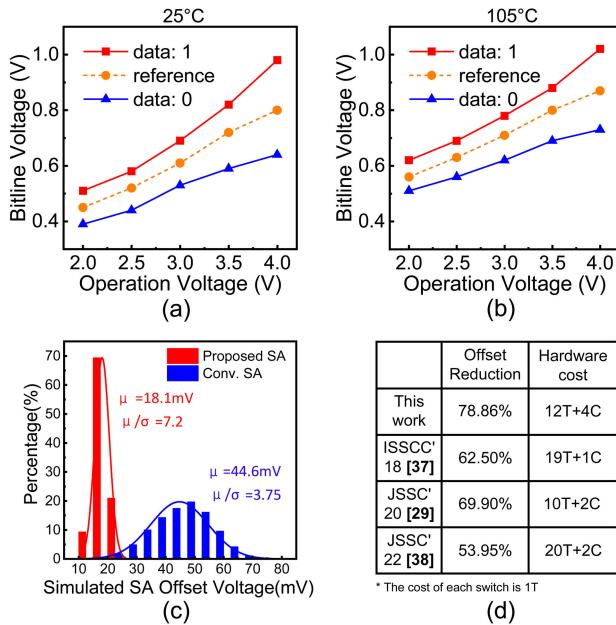


Fig. 6. (a) and (b) Measured reference voltage characteristics. (c) Offset voltage is reduced by $\sim 60\%$ according to 4k Monte Carlo simulations. (d) Comparison of the offset-canceled SA proposed in this work with the state-of-the-art works.



Fig. 7. Micrograph of the 9-Mb HZO-based FeRAM chip.

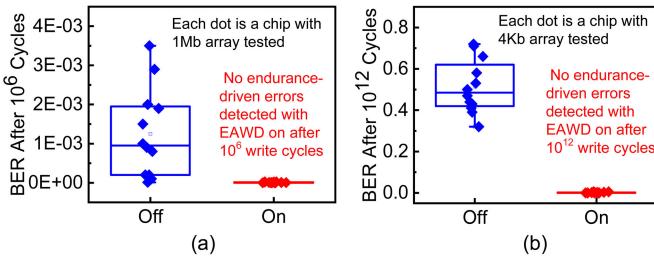


Fig. 8. Endurance test results. The BER of 12 chips is measured after 10^6 and 10^{12} cycles of write with and without the ECC-WD. (a) BER results after 10^6 cycles. (b) BER results after 10^{12} cycles.

To verify the reliability of the chip, the retention after cycles (RAC) characteristics are tested, as shown in Fig. 9(a) and (b). The samples are baked at 150°C for 1000 h, which is equivalent to 85°C for 10 years. Fig. 9(a) shows a BL voltage margin over 200 mV after a 10^6 write-cycle retention test, while Fig. 9(b) shows a BL voltage margin over 150 mV

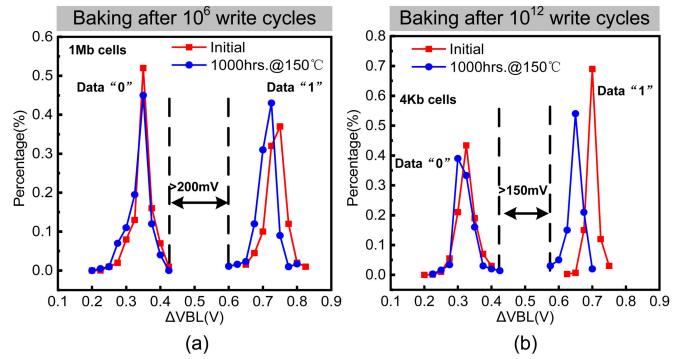


Fig. 9. Test results of RAC. The samples are baked at 150°C for 1000 h. The FeRAM macro achieves a robust read voltage window of 200 and 150 mV after 10^6 and 10^{12} RAC. (a) Read voltage window after 10^6 RAC. (b) Read voltage window after 10^{12} RAC.

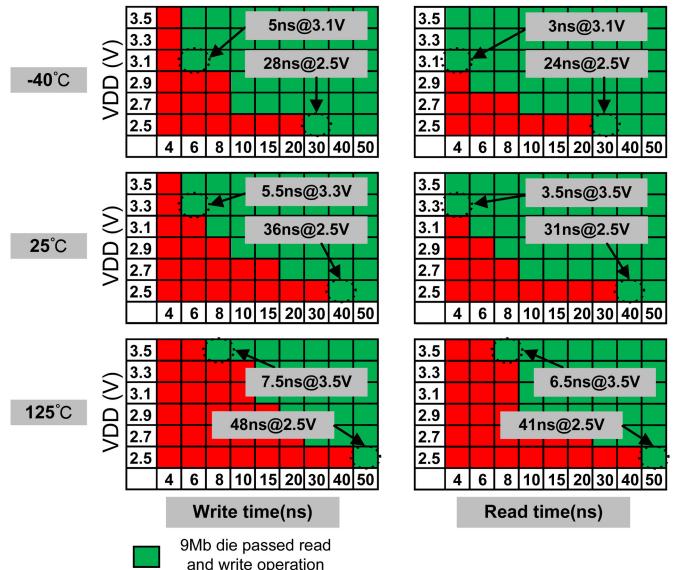


Fig. 10. Shmoo plots of the write and read operations at -40°C , 25°C , and 125°C .

after a 10^{12} write-cycle retention test. The result shows that the FeRAM macro achieves a robust read voltage window.

Fig. 10 shows a voltage-read time and voltage-write time from 5 to 40 ns. Shmoo plots with an operation voltage from 2.5 to 3.5 V are performed between -40°C and 125°C . All 9-Mb cells complete the read and write tests before the condition is deemed to pass. Measurement results show that the design achieves a 5-ns read sensing time and a 7-ns write time at 3.3 V and room temperature. At 3.1 V and -40°C , the FeRAM macro can achieve a 3-ns read time and a 5-ns write time. The improved read sense margin of SA makes it possible to reduce the chip operation voltage. The measured worst case read time and write time are 31 and 36 ns at 2.5 V and room temperature.

Table I compares this work with the state of the art. The proposed FeRAM chip has a 9-Mb memory capacity based on the 1T1C type. The retention time is 10 years/ 85°C , which is the best reported among the reported works of FeRAM. This work achieves the best write endurance and highest ferroelectric capacity among the reported work. Compared

with conventional FeRAM [8], this work is compatible with CMOS process and easy to scale.

V. CONCLUSION

This article introduces a 9-Mb 1T1C HZO-based FeRAM chip with high-performance write/read peripheral circuits. The write driver with an ECC-assist refresh scheme enables a high write voltage recovery process, which provides a circuit solution for the P_r degradation issue, and its temperature-aware capability improves the power efficiency by dynamic voltage tuning. The offset-canceled SA suppresses the negative effect of the offset voltage. Combining with the dummy-based reference generator for PVT variation compensation, this specially designed read circuit can efficiently improve the sense margin and reduce BER. By integrating these robust and high-performance peripheral circuits, we successfully implemented this mega-bit-level FeRAM test chip and proved the potential of embedded HZO-based FeRAM. The measurement results demonstrate the best write endurance and highest HZO-based ferroelectric storage capacity among the reported work as so far.

REFERENCES

- [1] M. M. Sharifi et al., "Application-driven design exploration for dense ferroelectric embedded non-volatile memories," in *Proc. IEEE/ACM Int. Symp. Low Power Electron. Design (ISLPED)*, Boston, MA, USA, Jul. 2021, pp. 1–6.
- [2] A. Agrawal, C. Wang, T. Sharma, and K. Roy, "Magnetoresistive circuits and systems: Embedded non-volatile memory to crossbar arrays," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 6, pp. 2281–2294, Jun. 2021.
- [3] J. Yang et al., "A 14 nm-FinFET 1 Mb embedded 1T1R RRAM with a 0.022 μm^2 cell size using self-adaptive delayed termination and multi-cell reference," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, vol. 64, Feb. 2021, pp. 336–338.
- [4] Y.-D. Chih et al., "A 22 nm 32 Mb embedded STT-MRAM with 10 ns read speed, 1M cycle write endurance, 10 years retention at 150 °C and high immunity to magnetic field interference," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2020, pp. 222–224.
- [5] Q. Dong et al., "A 1 Mb embedded NOR flash memory with 39 μW program power for mm-scale high-temperature sensor nodes," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2017, pp. 198–199.
- [6] T. Francois et al., "16 kbit HfO₂:Si-based 1T-1C FeRAM arrays demonstrating high performance operation and solder reflow compatibility," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2021, pp. 33.1.1–33.1.4.
- [7] J. Okuno et al., "SoC compatible 1T1C FeRAM memory array based on ferroelectric Hf_{0.5}Zr_{0.5}O₂," in *Proc. IEEE Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2020, pp. 1–2.
- [8] M. Qazi, M. Clinton, S. Bartling, and A. P. Chandrasekaran, "A low-voltage 1 Mb FeRAM in 0.13 μm CMOS featuring time-to-digital sensing for expanded operating margin in scaled CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2011, pp. 208–210.
- [9] S. Maeda et al., "A 20 ns-write 45 ns-read and 1014-cycle endurance memory module composed of 60 nm crystalline oxide semiconductor transistors," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2018, pp. 484–486.
- [10] M. Wei et al., "A scalable self-aligned contact NOR flash technology," in *Proc. IEEE Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2007, pp. 226–227.
- [11] H. Pon, "Technology scaling impact on NOR and NAND flash memories and their applications," in *Proc. 8th Int. Conf. Solid-State Integr. Circuit Technol.*, Shanghai, China, 2006, pp. 697–700.
- [12] G. Hu et al., "Key parameters affecting STT-MRAM switching efficiency and improved device performance of 400 °C-compatible p-MTJs," in *IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2017, pp. 38.3.1–38.3.4.
- [13] G. Hu et al., "STT-MRAM with double magnetic tunnel junctions," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2015, pp. 26.3.1–26.3.4.
- [14] A. M. S. Tossen, S. Yu, M. H. Anis, and L. Wei, "A study of the effect of RRAM reliability soft errors on the performance of RRAM-based neuromorphic systems," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 11, pp. 3125–3137, Nov. 2017.
- [15] M. Breitwisch et al., "Novel lithography-independent pore phase change memory," in *Proc. IEEE Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2007, pp. 100–101.
- [16] H. Y. Cheng et al., "Atomic-level engineering of phase change material for novel fast-switching and high-endurance PCM for storage class memory application," in *IEDM Tech. Dig.*, Washington, DC, USA, Dec. 2013, pp. 30.6.1–30.6.4.
- [17] H. L. Lung et al., "A novel low power phase change memory using inter-granular switching," in *Proc. IEEE Symp. VLSI Technol.*, Honolulu, HI, USA, Jun. 2016, pp. 1–2.
- [18] T. Mikolajick, U. Schroeder, and S. Slesazeck, "The past, the present, and the future of ferroelectric memories," *IEEE Trans. Electron Devices*, vol. 67, no. 4, pp. 1434–1443, Apr. 2020.
- [19] D. Takashima, "Overview of FeRAMs: Trends and perspectives," in *Proc. 11th Annu. Non-Volatile Memory Technol. Symp. Proc.*, Shanghai, China, Nov. 2011, pp. 1–6.
- [20] T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide thin films," *Appl. Phys. Lett.*, vol. 99, no. 10, pp. 1–12, Sep. 2011.
- [21] S.-C. Chang et al., "Anti-ferroelectric Hf_xZr_{1-x}O₂ capacitors for high-density 3-D embedded-DRAM," in *IEDM Tech. Dig.*, Dec. 2020, pp. 28.1.1–28.1.4.
- [22] T. Francois et al., "Demonstration of BEOL-compatible ferroelectric Hf_{0.5}Zr_{0.5}O₂ scaled FeRAM co-integrated with 130 nm CMOS for embedded NVM applications," in *IEDM Tech. Dig.*, Dec. 2019, pp. 15.7.1–15.7.4.
- [23] R. Cao et al., "Effects of capping electrode on ferroelectric properties of Hf_{0.5}Zr_{0.5}O₂ thin films," *IEEE Electron Device Lett.*, vol. 39, no. 8, pp. 1207–1210, Aug. 2018.
- [24] Z. Dang et al., "Improved endurance of Hf_{0.5}Zr_{0.5}O₂-based ferroelectric capacitor through optimizing the Ti–N ratio in TiN electrode," *IEEE Electron Device Lett.*, vol. 43, no. 4, pp. 561–564, Apr. 2022.
- [25] Y. Wang et al., "Precrystallization engineering of Hf_{0.5}Zr_{0.5}O₂ film in back-end-of-line compatible ferroelectric device for enhanced remnant polarization and endurance," *IEEE Electron Device Lett.*, vol. 44, no. 3, pp. 396–399, Mar. 2023.
- [26] J. Okuno et al., "1T1C FeRAM memory array based on ferroelectric HZO with capacitor under bitline," *IEEE J. Electron Devices Soc.*, vol. 10, pp. 29–34, 2022.
- [27] M. H. Park et al., "Giant negative electrocaloric effects in Hf_{0.5}Zr_{0.5}O₂ thin films," *Adv. Mater.*, vol. 28, no. 36, pp. 7956–7961, Sep. 2016.
- [28] B. Liu, J. Cai, J. Yuan, and Y. Hei, "A low-voltage SRAM sense amplifier with offset cancelling using digitized multiple body biasing," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 4, pp. 442–446, Apr. 2017.
- [29] J. M. Yoon, H. Do, D. Koh, S. H. Oak, J. Lee, and D.-K. Jeong, "A capacitor-coupled offset-canceled sense amplifier for DRAMs with reduced variation of decision threshold voltage," *IEEE J. Solid-State Circuits*, vol. 55, no. 8, pp. 2219–2227, Aug. 2020.
- [30] M. Sharifkhani, E. Rahiminejad, S. M. Jahanuzzaman, and M. Sachdev, "A compact hybrid current/voltage sense amplifier with offset cancellation for high-speed SRAMs," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 5, pp. 883–894, May 2011.
- [31] W. Zhao, C. Chappert, V. Javerliac, and J.-P. Noziere, "High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits," *IEEE Trans. Magn.*, vol. 45, no. 10, pp. 3784–3787, Oct. 2009.
- [32] J. Yang et al., "A 9 Mb HZO-based embedded FeRAM with 1012-cycle endurance and 5/7 ns read/write using ECC-assisted data refresh and offset-canceled sense amplifier," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2023, pp. 1–3.
- [33] T. Gong et al., "10⁵× endurance improvement of FE-HZO by an innovative rejuvenation method for 1z node NV-DRAM applications," in *Proc. Symp. VLSI Technol.*, Kyoto, Japan, Jun. 2021, pp. 1–2.
- [34] S. J. Kim, J. Mohan, S. R. Summerfelt, and J. Kim, "Ferroelectric Hf_{0.5}Zr_{0.5}O₂ thin films: A review of recent advances," *JOM*, vol. 71, no. 1, pp. 246–255, Jan. 2019.

- [35] R. Micheloni et al., "A 4Gb 2b/cell NAND flash memory with embedded 5b BCH ECC for 36 MB/s system read throughput," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, 2006, pp. 497–506.
- [36] R. Ogiwara, D. Takashima, S. Doumae, S. Shiratake, R. Takizawa, and H. Shiga, "Highly reliable reference bitline bias designs for 64 Mb and 128 Mb chain FeRAMs," *IEEE J. Solid-State Circuits*, vol. 50, no. 5, pp. 1324–1331, May 2015.
- [37] Q. Dong et al., "A 1 Mb 28 nm STT-MRAM with 2.8 ns read access time at 1.2 V VDD using single-cap offset-cancelled sense amplifier and in-situ self-write-termination," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, San Francisco, CA, USA, Feb. 2018, pp. 480–482.
- [38] M. Nakano et al., "A 40-nm embedded SG-MONOS flash macro for high-end MCU achieving 200-MHz random read operation and 7.91-Mb/mm² density with charge-assisted offset cancellation sense amplifier," *IEEE J. Solid-State Circuits*, vol. 57, no. 10, pp. 3094–3102, Oct. 2022.
- [39] K. Hoya et al., "A 64-Mb chain FeRAM with quad BL architecture and 200 MB/s burst mode," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 12, pp. 1745–1752, Dec. 2010.
- [40] H. Shiga et al., "A 1.6 GB/s DDR2 128 Mb chain FeRAM with scalable octal bitline and sensing schemes," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 142–152, Jan. 2010.



Qiqiao Wu (Student Member, IEEE) received the B.S. and M.Eng. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2019 and 2022, respectively. He is currently pursuing the Ph.D. degree with the Frontier Institute of Chip and System, Fudan University, Shanghai, China.

His current research interests include memory circuit design, security, and stochastic computing based on nonvolatile memory device.



Yue Cao received the B.S. degree in electrical and computer engineering from Shanghai Jiaotong University, Shanghai, China, in 2015, and the M.Eng. degree in electrical engineering from the University of California, Berkeley, CA, USA, in 2016. She is currently pursuing the Ph.D. degree with Fudan University, Shanghai.

She worked at Intel, Santa Clara, CA, USA, from 2016 to 2020, and then at the Zhejiang Laboratory, Hangzhou, China, from 2021 to 2023, as an SOC Design Engineer. She has authored four issued patents in China with one pending. Her research interests include nonvolatile memory and processing-in-memory circuit design.



Qing Luo (Senior Member, IEEE) received the Ph.D. degree in microelectronics from the University of Chinese Academy of Sciences, Beijing, China, in 2017.

In 2017, he joined the Institute of Microelectronics of Chinese Academy of Sciences, Beijing, as an Associate Professor, where he currently serves as a Full Professor. His research interests include hafnia-based ferroelectric memory and 3-D integration techniques of RRAM.



Haijun Jiang received the B.S. degree in automation from Henan Polytechnic University, Jiaozuo, China, in 2012, and the M.Eng. degree in software engineering from Central South University, Changsha, China, in 2015.

He currently works at Zhangjiang Laboratory, Shanghai, China, as an IC Test Engineer. He is engaging in research on testing field of new architecture memory chips.



Zhongze Han received the B.S. degree in microelectronics from the Dalian University of Technology, Dalian, China, in 2020. He is currently pursuing the Ph.D. degree with the Key Laboratory of Microelectronics Devices and Integrated Technology, Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China.

His current research interests include memory circuit design and 3-D heterogeneous integrated near-memory computing.



Yongkang Han received the M.Eng. degree in integrated circuit engineering from Fudan University, Shanghai, China, in 2021.

He currently works at Zhangjiang Laboratory, Shanghai, as an Analog IC Design Engineer. He is responsible for new architecture memory research and development.



Chunmeng Dou (Member, IEEE) received the B.S. degree in physics from Nanjing University, Nanjing, China, in 2009, and the M.S. and Ph.D. degrees in electronics and applied physics from the Tokyo Institute of Technology, Tokyo, Japan, in 2011 and 2014, respectively.

From 2014 to 2016, he was a Research Associate with the University of Cambridge, Cambridge, U.K., working on resistive memory. From 2016 to 2017, he was with SMIC, Shanghai, China, working on embedded flash memory. From 2017 to 2018, he was

with National Tsing-Hua University, Hsinchu, Taiwan, working on computing in-memory. In 2018, he joined the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China, where he is currently a Professor. His research interests include emerging memory circuit design, design technology co-optimization, and emerging computing paradigms.



Hangbing Lv (Senior Member, IEEE) received the Ph.D. degree in microelectronics from Fudan University, Shanghai, China, in 2009.

In 2010, he joined the Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China, as a Post-Doctoral Research Fellow, where he currently serves as a Full Professor. He has more than 15 years of experience on research and development on emerging nonvolatile memories, including device physics, integration, macrodesign, and applications. His current research interests include resistive switching memory, hafnia-based ferroelectric memory, and new computing paradigm.



Jianguo Yang (Member, IEEE) received the Ph.D. degree in microelectronics from Fudan University, Shanghai, China, in 2016.

In 2016, he joined the Department of Microelectronics, Fudan University, as a Post-Doctoral Research Fellow. In 2019, he joined the Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China, as an Associate Professor. He is currently a Professor with Zhangjiang Laboratory, Shanghai. His research interests include memory circuit design, hardware security, and new computing paradigm.



Qi Liu (Member, IEEE) received the Ph.D. degree from Anhui University, Anhui, China, in 2010.

He joined the Institute of Microelectronics of the Chinese Academy of Sciences (IMECAS), Beijing, China, as a Professor, in 2016. He is currently a Professor with the Frontier Institute of Chip and System, State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai, China, and a Guest Professor with the IMECAS. His research interests focus on fabrication, characterization, and mechanism of emerging memory devices and chips for memory, logic, computing-in-memory, and neuromorphic computing.



Ming Liu (Fellow, IEEE) received the Ph.D. degree in material engineering from the Beijing University of Aeronautics and Astronauts, Beijing, China, in 1998.

She joined the Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, in 1999, where she became a Full Professor in 2000. She was selected as an Academician of the Chinese Academy of Sciences, Beijing, in 2015. She was the Director of the Key Laboratory of Microelectronics Devices and Integrated Technology, Institute of Microelectronics of the Chinese Academy of Sciences. Currently, she is the Director with the Frontier Institute of Chip and System, State Key Laboratory of Integrated Chips and Systems, Fudan University, Shanghai, China. Her research areas include advanced semiconductor process, emerging memory and logic technologies, novel integration and package technology, and next-generation computing paradigms.

Dr. Liu received the National Natural Science Award (Second Level) once, the National Invention Award (Second Level) three times, the National Science and Technology Progress Award (Second Level) once, the Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, and the Achievement Prize of Chinese Vacuum Science and Technology (Leybold Prize).