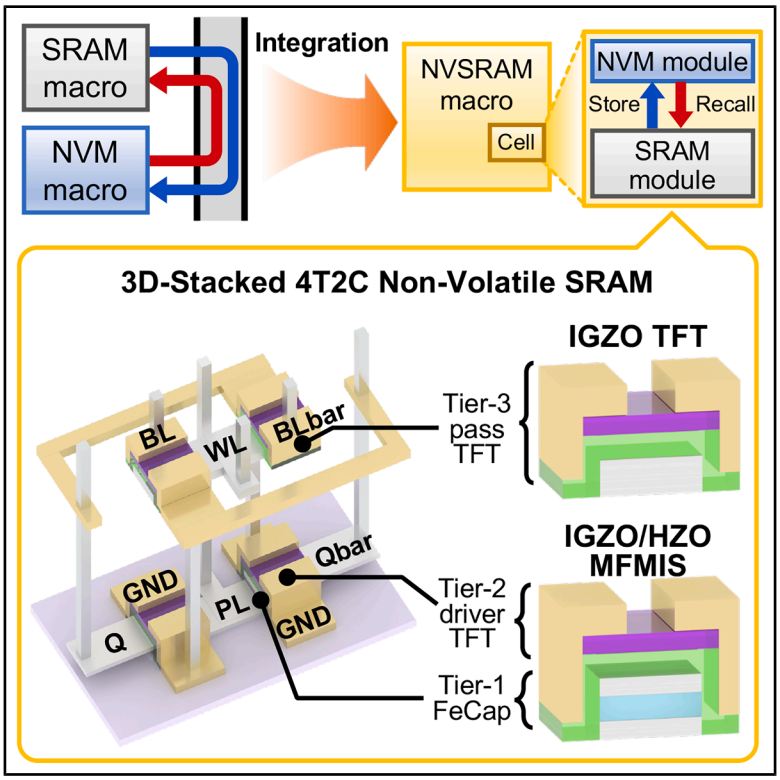
Device

Article

High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors

Graphical abstract



Highlights

- Monolithic 3D-integrated SRAM with smaller footprint and nonvolatile data retention
- 3 tiers of IGZO TFTs and hafnia-based ferroelectric memories are vertically stacked
- M3D-stacked nvSRAM exhibits enhanced density, speed, endurance, and power efficiency

Authors

Qihan Liu, Yu Li, Liwei Liu, ..., Peng Lin, Yingfen Wei, Qi Liu

Correspondence

haoj@fudan.edu.cn (H.J.), yingfen_wei@fudan.edu.cn (Y.W.), qi_liu@fudan.edu.cn (Q.L.)

In brief

Liu et al. introduce a memory design that combines emerging thin-film transistors and advanced ferroelectric memories through a monolithic 3D integration method. The fabrication procedures are fully BEOL compatible under a low processing temperature of ≤400°C. A high-performance and nonvolatile SRAM is demonstrated, featuring a reduced two-transistor cell size, robust data retention for 10^5 s at 85°C, 0.28/0.45/0.26 V hold/write/read noise margins at a low 0.9 V supply voltage, and ultra-low static power.

2 Explore
Early prototypes with exciting performance and new methodology

Liu et al., 2025, Device 3, 100833
September 19, 2025
Published by Elsevier Inc.
https://doi.org/10.1016/j.device.2025.100833

CellPress

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

Device

CellPress
OPEN ACCESS

Article

High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors

Qihan Liu,[1,6] Yu Li,[1,6] Liwei Liu,[1] Hao Jiang,[1,7,*] Haozhe Zhu,[1] Zexing Chen,[1] Yang Yang,[2] Mengwei Si,[3] Can Li,[4] Peng Lin,[5] Yingfen Wei,[1,*] and Qi Liu[1,*]

[1]State Key Laboratory of Integrated Chips and Systems, Frontier Institute of Chip and System, Fudan University, Shanghai, China
[2]Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China
[3]Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
[4]Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China
[5]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[6]These authors contributed equally
[7]Lead contact
*Correspondence: haoj@fudan.edu.cn (H.J.), yingfen_wei@fudan.edu.cn (Y.W.), qi_liu@fudan.edu.cn (Q.L.)
https://doi.org/10.1016/j.device.2025.100833

THE BIGGER PICTURE Static random-access memory (SRAM) distinguishes itself from other memory technologies with its ultrafast access speed and robust cycling endurance under low operating voltage. These advantages have made SRAM indispensable in high-performance applications such as CPU caches and embedded memory in FPGAs, where speed and reliability are paramount. However, its widespread adoption is constrained by two major drawbacks: a large cell size resulting from the excessive use of transistors per data bit and considerable power consumption caused by persistent leakage currents to preserve data. To address these limitations, we developed a CMOS-compatible integration technique that vertically stacks logic transistors and memory components with more compact and complex interconnect structures. This approach enables increased device density, reduced interconnect lengths with enhanced bandwidth and energy efficiency, and extended capabilities by incorporating multifunctional devices.

SUMMARY

Monolithic three-dimensional (M3D) integration of back-end-of-line (BEOL)-compatible and high-performance transistors and memory devices offer a promising strategy to overcome the limitations of conventional silicon-based computing techniques. Specifically, static random-access memory (SRAM), while valued for speed and endurance, faces the challenges of a large footprint and static power consumption. Here, we present a four-transistor-two-capacitor (4T2C) nonvolatile SRAM (nvSRAM) utilizing vertically stacked indium gallium zinc oxide (IGZO) thin-film transistors (TFTs) and $Hf_{0.5}Zr_{0.5}O_2$ (HZO)-based ferroelectric capacitors (FeCaps) under a BEOL-compatible thermal budget. The HZO FeCaps and IGZO TFTs are optimized toward their M3D integration with high performance and inter-tier uniformity. Our 3-tier M3D-integrated 4T2C nvSRAM achieves a 2-transistor (2T) footprint and data retention at 85°C for 10^5 s after power off. Moreover, design space and optimization strategies of the 4T2C nvSRAM are explored for higher operation speed, lower energy consumption, and more robust data stability.

INTRODUCTION

The back-end-of-line (BEOL) process connects individual devices such as transistors, capacitors, and resistors fabricated at the front end of line through depositing and patterning metal interconnect layers, which is critical for silicon-chip fabrication. BEOL-compatible and high-performance oxide-semiconductor (OS)/two-dimensional (2D) materials/carbon nanotubes (CNT)-based transistors[1–8] and emerging memory devices that include resistive and ferroelectric random-access memory (RRAM/FeRAM)[9–15] have triggered significant interest in their integration into three-dimensional (3D) multi-tier circuits. Such monolithic 3D (M3D) integration that vertically stacks various functional layers onto the same chip (Figure 1A) can increase the density of devices, shorten interconnect lengths with enhanced bandwidth and energy efficiency, and enable multifunctionalities

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

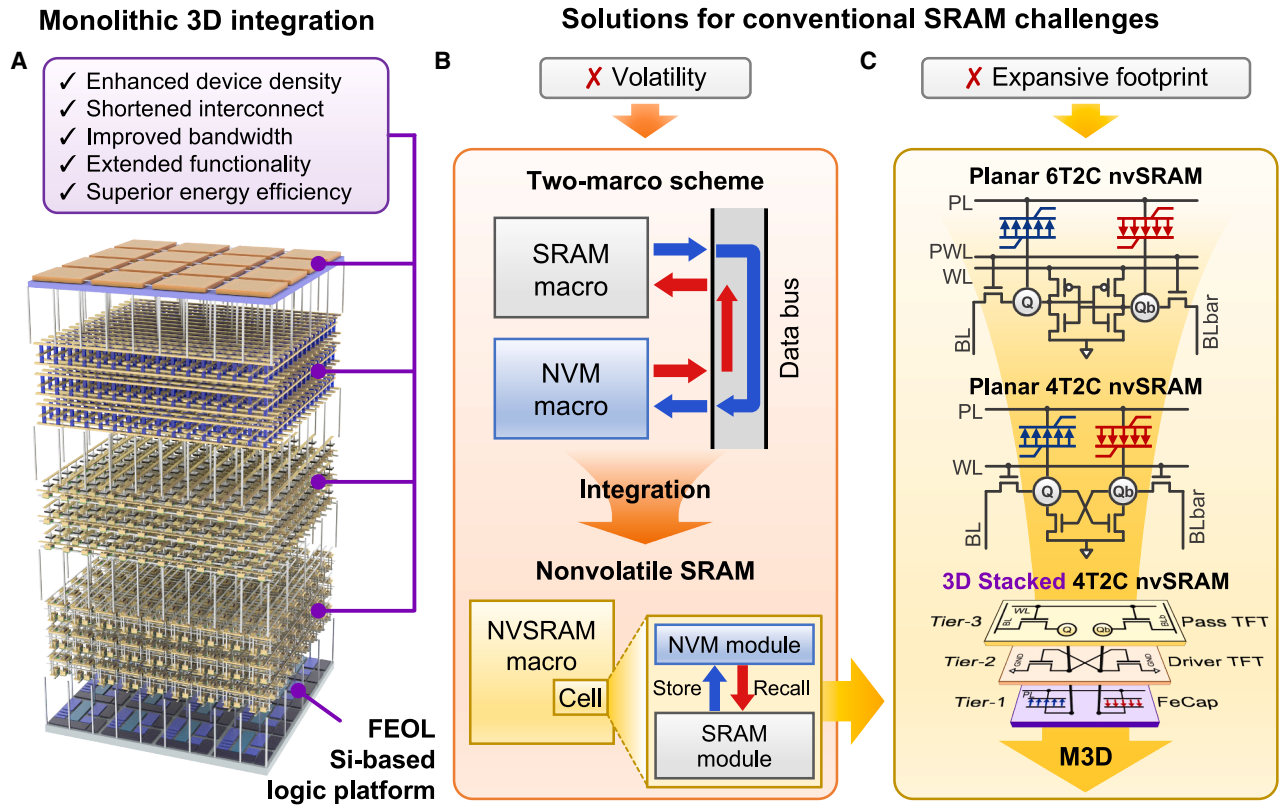


Figure 1. Illustration of M3D integration for high-density nvSRAM

(A) M3D integration of BEOL-compatible devices enables enhanced device density, shortened interconnect, improved bandwidth, and extended functionality with superior energy efficiency.

(B and C) Solutions to overcome challenges of conventional SRAM in its volatility and expansive footprint.

(B) Volatility-related issues can be effectively addressed by the implementation of nvSRAM through integrating nonvolatile modules onto SRAM.

(C) Innovations in nvSRAM structures from planar 6T2C and planar 4T2C to M3D-stacked 4T2C with significant area efficiency.

through incorporating devices for sensing, radiofrequency, hardware security, etc., providing a feasible pathway toward "more Moore" and "more than Moore."[16–20]

Static RAM (SRAM) shows high speed and enhanced endurance of read/write cycles with a low operation voltage (V_{dd}) and has been used as a crucial memory component in various hardware systems.[21,22] However, SRAM technology faces drawbacks due to its volatile nature and the large cell size of a typical six-transistor (6T) design (Figures 1B and 1C). The former results in static power consumption dominated by the leakage current of transistors, which becomes more significant at advanced Si-CMOS nodes. Off-chip nonvolatile memory, e.g., embedded Flash, is usually required to back up data before powering off SRAM for low-power artificial intelligence of things (AIoT) applications (e.g., mobile devices).[23] The two-macro scheme (Figure 1B) increases the costs and causes extra delay and energy when powering on/off. By integrating nonvolatile modules onto the volatile SRAM cells (Figure 1B), nonvolatile SRAM (nvSRAM) becomes one of the prevalent solutions for overcoming the above challenges. In this way, when powering on/off is necessary, data can be transferred between nonvolatile modules and SRAM cells in a parallel bit-to-bit manner through the so-called recall/store operation.

Various nvSRAM architectures have been proposed, including RRAM-based 4T-two-resistor (4T2R)[24] and 7T1R,[25] magnetic tunnel junction (MTJ)-based 4T-two-MTJ (4T2M),[26] ferroelectric FET (FeFET)-based 8T,[27] phase-change memory (PCM)-based 8T2R,[28] charge trapping (CT)-based 8T,[29] and ferroelectric capacitor (FeCap)-based 6T-two-capacitor (6T2C).[23,30] Compared to these reported implementations, FeCap-based 6T2C nvSRAM differs by using two FeCaps directly tied to the storage nodes of a conventional 6T SRAM structure (Figure 1C), and such a simplification becomes practical after the discovery of ferroelectricity in CMOS-compatible hafnia-based thin films.[31] Specifically, the utilization of FeCaps eliminates additional access transistors to suppress the direct current (DC)-short current of the nonvolatile elements and does not require intricate control schemes and peripheral circuits for data backup and restoration. Moreover, the low power consumption, robust endurance, and retention of the FeCaps ensure low energy consumption and a long lifetime for frequent data recall and storage. Recently, the potential of such 6T2C nvSRAM has been demonstrated and evaluated at both the device and chip levels.[30,32,33]

Although nvSRAM mitigates the volatility-related issue of SRAM, the challenge in density due to the large cell footprint remains, limiting capacity and further development. The emerging M3D

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

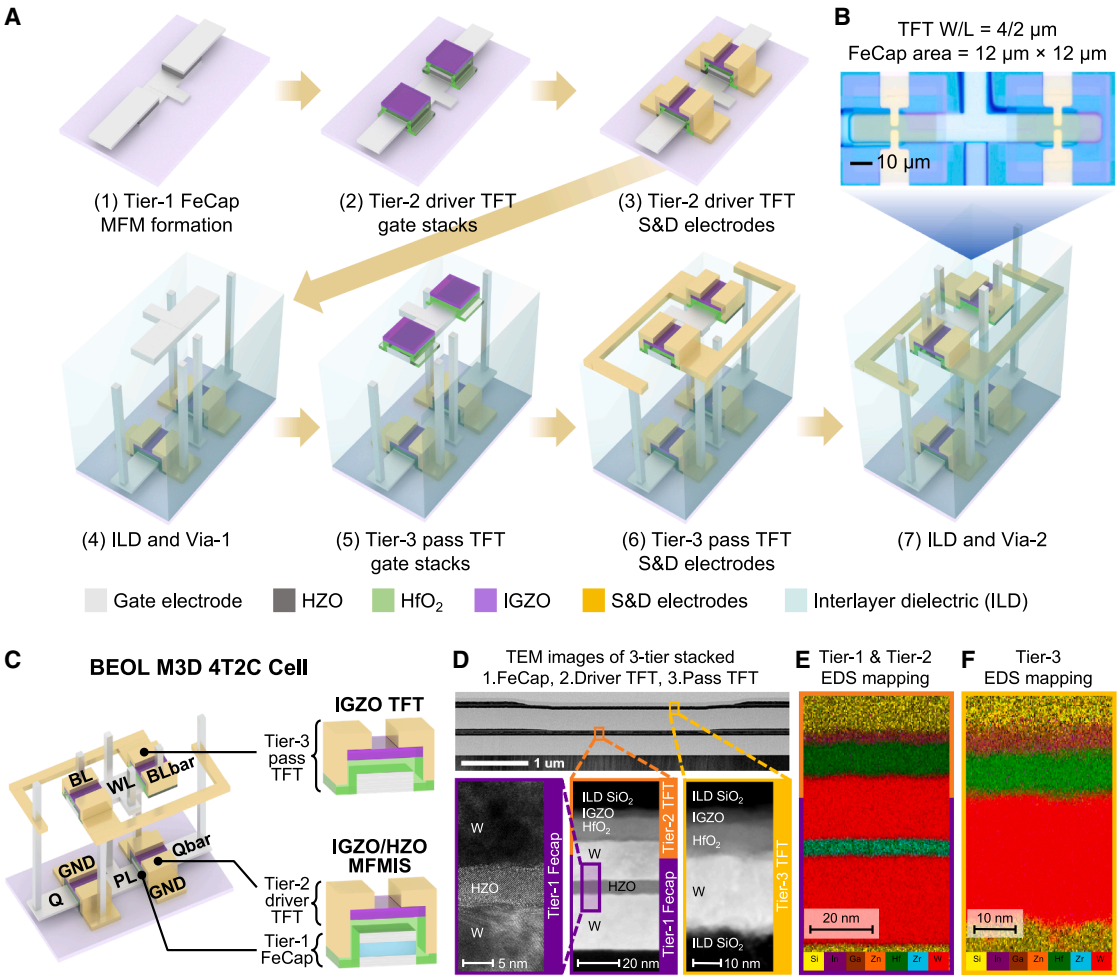


Figure 2. Fabrication and physical characterization of M3D-4T2C nvSRAM

(A) Schematic diagram for the fabrication process flow of the M3D-4T2C cell. S&D, source and drain of TFTs.

(B) Top-view optical image of a fabricated 4T2C cell.

(C) Magnified schematic showing the M3D-4T2C cell with stacked FeCaps and IGZO TFTs. Plate line (PL), word line (WL), bit lines (BL/BLbar), storage nodes (Q/Qbar), and the ground (GND) are labeled.

(D–F) Cross-sectional HRTEM and HAADF-STEM images (D) with EDS mappings of vertically stacked FeCaps in tier 1, driver TFTs in tier 2, and pass TFTs in tier 3 (E and F).

integration offers an opportunity to tackle this issue (Figure 1C). In this work, we demonstrate M3D-integrated nvSRAM featuring vertically stacked indium gallium zinc oxide (IGZO) thin-film transistors (TFTs) and $Hf_{0.5}Zr_{0.5}O_2$ (HZO)-based FeCaps with a low processing temperature of $\leq$400°C. This design achieves a small 2T footprint, which can ideally approach a ~67% reduction in area (Figure 1C). The 4T2C structure is used here to save the transistor counts, as well as reduce the fabrication complexity and area size compared to the 6T2C configuration. IGZO TFTs are adopted for their low power consumption, high electron mobility, and BEOL compatibility.[13,34,35] IGZO TFTs are more mature in manufacturing when compared with other 2D/CNT-based competitors, as they have been commonly used in displays.[36] The electrical performance and fabrication process are optimized for HZO FeCaps and IGZO TFTs toward their M3D integration. The 3D-stacked 4T2C nvSRAMs are fabricated, with their functionality demonstrated and discussed. The stored data can be recalled even after 10^5 s at 85°C, and a hold noise margin (HNM) of 0.28 V/write noise margin (WNM) of 0.45 V/read noise margin (RNM) of 0.26 V are achieved with a low V_{dd} of 0.9 V, illustrating the nonvolatility and reliability of our M3D nvSRAM. We evaluate our 4T2C nvSRAM through extensive simulations and propose the design space for their further optimizations. Our proof-of-concept work demonstrates a solution for memory technology with advantages in density, speed, endurance, power consumption, and compatibility with M3D integration.

RESULTS AND DISCUSSION

M3D integration of 2T-footprint nvSRAM

Figure 2A illustrates schematics of the fabrication process flow of the M3D-integrated 4T2C nvSRAM, with the key fabrication

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

steps shown in methods, Note S1, and Figure S1. The fabrication temperature is ≤400°C throughout the whole process. 30 nm W was used as top and bottom electrodes for HZO FeCaps. 6 nm IGZO and 10 nm HfO_2 were used as the channel and gate dielectric layers, respectively, for both driver and pass TFTs. For the vertical stacking of multi-tiers, 300 nm SiO_2 is deposited by plasma-enhanced chemical vapor deposition (PECVD) as the interlayer dielectric (ILD), which can also passivate the IGZO channel, according to previous studies.[37–39] The top-view optical image of the fabricated cell is presented in Figure 2B. The magnified schematic in Figure 2C shows the 3-tier sequential stacking of FeCaps, driver TFTs, and pass TFTs, as well as the plate line (PL), word line (WL), bit-line pair (BL and BLbar), storage node pair (Q and Qbar), and the ground (GND). As a result, the effective cell size can be reduced, approaching the footprint of a 2T configuration. The cross-sectional transmission electron microscopy (TEM) and high-angle annular dark-field (HAADF) scanning TEM (STEM) images provide a detailed view of the three stacked tiers (Figure 2D). Figures 2E and 2F show the energy-dispersive X-ray spectroscopy (EDS) mapping profiles of our M3D-4T2C nvSRAM device, capturing the elemental distributions of the W/HZO/W FeCap at tier 1, IGZO TFTs at tier 2 and tier 3, and the SiO_2 ILD. The distinct signals from different elements verify clean interfaces, negligible inter-tier diffusions, and uniformly deposited stacks. The top-view scanning electron microscopy (SEM) image of the fabricated 4T2C device, along with more cross-sectional SEM and TEM images, is illustrated in Figure S2.

Optimization of HZO FeCaps and IGZO TFTs for M3D integration

For FeCaps, we chose tungsten (W) to act as the electrodes. It has been reported that W has a lower coefficient of thermal expansion (CTE) when compared with other commonly used electrodes such as TiN, contributing to higher tensile stress and hence more enhanced ferroelectricity in HZO.[40] Moreover, robust retention under temperatures up to 125°C and excellent endurance of 10^{12} cycles were demonstrated in W/HZO/W according to previous work.[41] 7 nm, instead of regular 10 nm, HZO was used in our study for a lower operation voltage. The W/7 nm HZO/W capacitor requires only a low thermal budget (400°C, 10 min) for the formation of the ferroelectric orthorhombic- (o-)phase within the HZO layer (Figure S9). The high-resolution TEM (HRTEM) and the corresponding fast Fourier transform (FFT) images for the HZO layer, shown in Figure S10, suggest a crystalline structure and the presence of the o-phase. Figure 3A shows the polarization-voltage (*P-V*) and current density-voltage (*J-V*) loops from a woken-up FeCap (area size: 50 × 50 μm) with 10-kHz triangle pulses, where a high double remnant polarization ($2P_r$) of ~45 μC/cm^2 was observed. The W/HZO/W FeCap showed a $2P_r$ of 35 μC/cm^2 even in its pristine state, and the *P-V* loop became almost saturated after 10 wake-up cycles (Figure S11). The leakage current of our FeCap was characterized by the DC measurement, and its conduction mechanism can be attributed to the Poole-Frenkel emission (Figure 3B; see methods for details), similar to that reported by Park et al.[42] The FeCap leakage is a critical parameter for the functioning of the 4T2C nvSRAM, which will be discussed in detail in a following section. The fabrication of IGZO TFTs on both upper tiers demands 10 min rapid thermal processing (RTP) at 300°C in O_2 to reduce oxygen vacancies in IGZO channels.[43] Hence, we compared the *P-V* behaviors from our FeCaps with and without the extra thermal treatment, where no obvious change was observed (Figure 3C), confirming the W/HZO/W FeCap is friendly with the following M3D TFT integration.

We examined the performance of our IGZO TFTs. Figure 3D presents the measured drain current (I_d)-gate voltage (V_g) transfer curves of the fabricated IGZO TFT (channel width-to-length ratio [*W/L*] of 10:2 μm) with the source grounded and the drain voltage (V_d) varied from 0.1 to 3 V. The gate leakage currents were negligibly low throughout all measurements. The device exhibits a positive threshold voltage (V_{th}) of ~0.7 V, a saturation mobility of 15.8 cm^2/Vs, a subthreshold swing (*S.S.*) of 125 mV/decade, an on-current exceeding 40 μA/μm, and an on/off current ratio greater than 10^8, compatible with the recent report.[1] The ILD layer plays a crucial role in isolating different tiers for M3D integration.[6,39,44–46] The I_d-V_g curves of IGZO TFTs with and without a PECVD-SiO_2 ILD layer on top are compared in Figure 3E. The deposition of the ILD induces an ~1 V negative shift in V_{th}, which is attributed to hydrogen incorporation into the IGZO channel.[38] The hydrogen can act as a donor, increase the carrier concentration in the channel, and thereby reduce the V_{th}. However, TFTs with a positive V_{th} are preferred for the proper functioning of the 4T2C nvSRAM. As a result, further optimizations of the IGZO TFT fabrication processes are necessary. To obtain a positive V_{th} to ensure stable SRAM operations, we investigated V_{th} modulation by optimizing the IGZO deposition process. The IGZO TFT V_{th} could be modulated by reducing the sputtering power (Figure 3F).[47] After reducing the sputtering power while maintaining a roughly similar IGZO thickness and In:Ga:Zn stoichiometry, a positive V_{th} shift was achieved, which has been attributed to a lower oxygen vacancy concentration.[47,48] We explored further shifting the V_{th} positively by decreasing the IGZO channel thickness by controlling the deposition time (Figures 3F and S12). It has been reported that the lower carrier concentration[49] and lower number of intrinsic free charge carriers[50] resulting from thinner channel films lead to the observed positive V_{th} shift. We investigated the impact of the ILD SiO_2 deposition conditions on the V_{th}. Figure 3G illustrates that decreasing the PECVD power during SiO_2 deposition further increases V_{th}, which can be explained by a lower PECVD power reducing the plasma energy and thereby minimizing the possible hydrogen-doping effect in the IGZO layer.[51] Dynamic secondary ion mass spectrometry (D-SIMS) measurements confirm that the hydrogen concentration in the IGZO layer decreases with reduced ILD deposition power (Figure S13), strongly supporting our above hypothesis.

With these optimized fabrication processes, we sequentially stacked the tier-2 and tier-3 IGZO TFTs onto tier-1 FeCaps. Figures 3H and 3I compare the statistical results of transfer characteristics of TFTs from both tiers based on measurements from 40 devices in each tier (inset in Figure 3H). The results demonstrate nearly consistent performance between different tiers, with median V_{th}/*S.S.* values of 0.75 V/160 mV/decade for tier 2 and 0.71 V/126 mV/decade for tier 3, which underscores the potential of our optimized fabrication process flow of IGZO

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

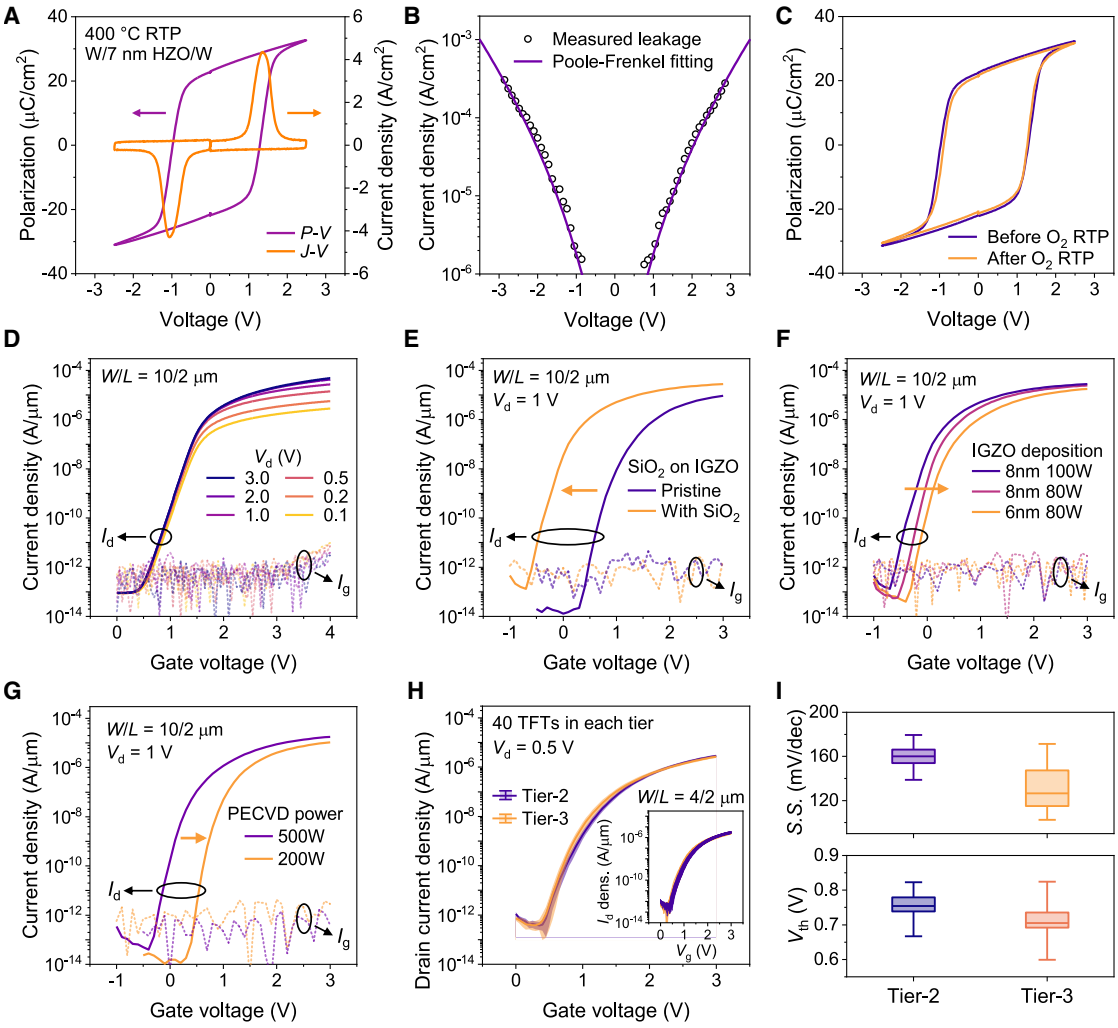


Figure 3. Electrical behavior and optimization of HZO-based FeCaps and IGZO TFTs

(A and B) Measured *P-V* and *J-V* loops of ferroelectric switching (A) with static leakage current density (B) in 7 nm HZO-based FeCaps.

(C) Measured *P-V* loop of HZO-based FeCap (50 × 50 μm) before/after RTP O_2 annealing process. The *P-V* loops are measured with 10 kHz triangular pulses.

(D) Measured transfer characteristics of IGZO TFTs.

(E) SiO_2 ILD on top of IGZO TFTs leads to an ~1 V negative shift in its V_{th}.

(F and G) Optimizations in V_{th} by lower IGZO deposition power and thinner IGZO thickness (F) and lower SiO_2 ILD deposition power (G).

(H) Device-to-device variation of tier-2 and tier-3 IGZO TFTs from 40 devices (80 TFTs in total). Inset: transfer curves from each tier (V_d = 0.5 V).

(I) *S.S.* and V_{th} distributions in tier-2 and tier-3 TFTs extracted from (H).

Solid/dashed lines in (D) and (G): the drain/gate leakage current densities (I_d/I_g). Thick orange arrows indicate the V_{th} shift direction.

TFTs for M3D integration. It should be noted that TFTs in tier 2 have a smaller variation compared with those in tier 3. Possible mechanisms, including surface roughness of the channel and gate bump roughness, have been proposed by Yuvaraja et al.[1] to explain the differences in variations in TFT electrical performances across multi-stacked layers. The higher thermal budget experienced by the tier-2 TFTs may also play a role in its improved variation, as the thermal budget is critical for the IGZO TFT performance. Likewise, the higher thermal budget experienced by the tier-2 TFTs may also play a role in its improved variation, as the thermal budget is critical for the IGZO TFT performance.[52]

Functionality demonstration of M3D-4T2C nvSRAM

Building upon the optimized FeCaps and IGZO TFTs developed in the previous section, we constructed a 4T2C nvSRAM cell through M3D integration. The optical image of an array of nvSRAM cells is illustrated in Figure S14. The working principles and timing chart for read and write operations in normal SRAM mode, as well as storing (store) and retrieving (recall) nonvolatile data through the FeCaps, are detailed in Note S2 and Figure S3.[53] Our proposed circuit design for an integrated M3D-4T2C nvSRAM array is shown in Figure S15, which allows the store and recall operations to be performed with minimal adjustments to a conventional 6T SRAM circuit, thanks to the

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

compatibility and flexibility offered by our optimized device structures.

Figure 4A shows the WL/BL/BLbar waveforms and Figure 4B shows the measured voltages of Q/Qbar nodes during the read and write operations of our M3D nvSRAM in the normal SRAM mode with V_{dd} = 0.9 V. As shown in Figure 4A, data "0" were stored in the initial state with Q at GND and Qbar at V_{dd}. To verify the write "1" operation, BL was set to V_{dd}, and BLbar was grounded. As WL turned on, Q was brought up to V_{dd}, whereas Qbar was discharged from V_{dd} to GND through pass TFTs. Q/Qbar voltages (data 1) could be maintained when the WL/BL/BLbar voltages decreased to 0 V. The write 0 operation was sequentially performed by setting BL to 0 V and BLbar to V_{dd}, which alters Q to GND and Qbar to V_{dd} after turning on pass TFTs. We demonstrated that both 0 and 1 states could be held for a long time over 100 s, which should be attributed to the optimized balance between the static currents of IGZO driver TFTs and FeCaps and indicates the reliability of the fabricated M3D-4T2C nvSRAM. Due to large parasitic capacitance, the write operation is demonstrated in the millisecond range. To address this limitation, several optimization strategies are proposed to suppress parasitic effects, aiming for high-frequency operation and high-performance computational systems (Note S3).

The incorporation of FeCaps in the 4T2C structure enables the nonvolatility. The operation to save the nonvolatile data to FeCaps before power off is indicated as store, and on the contrary, the operation to retrieve the stored data during power-on is defined as recall. Our store operation testing was implemented as a proof-of-concept demonstration, focusing on confirming the polarization switching of two FeCaps during data storage (Figures 4C, 4D, and S16). The schematic for storing data 1 is shown in Figure 4C, and the operation details, including node voltage and FeCap switching, are illustrated by our simulation results in Figure S16.

For a proof-of-concept demonstration of the recall operation of data 1 (Figures 4E, 4F, and S17), two 470 pF capacitors are coupled to BL and BLbar nodes, respectively, acting as the BL parasitic capacitances in an SRAM array. The utilization of those capacitances ensures that PL voltage can be effectively applied on the FeCaps to switch the polarization. The stored data on FeCaps can be recalled by sensing the BL voltage difference (Figure 4F). Recent advancements have enabled the sensing of signal margins smaller than 0.1 V,[54] and the state-of-the-art comparator has achieved high sensitivity in detecting voltage differences as small as 10 mV,[55] ensuring the robustness of the recall operation. Moreover, for future chip-level applications with more specialized and advanced readout circuits, a higher sensing margin can be expected. Okuno et al. reported a sensing margin of >0.48 V with HZO-based FeCaps of similar $2P_r$ to ours.[56] Then, a write-back operation is conducted to write the recalled data into the device by setting nodes Q and Qbar to corresponding voltage levels (V_{dd}/0 V). In this manner, the nonvolatility of our 4T2C nvSRAM mainly depends on the retention property of the FeCaps after power removal. Extensive studies have reported robust 10 year retention in HZO-based FeCap devices[57–59] and arrays,[60,61] and pronounced retention at 150°C over 10^6 s has been shown in a 128 kb 1T1C FRAM chip.[61] As shown in Figure 4F, the stored data 1 can be successfully recalled even after the M3D-4T2C device is baked at 85°C for 10^5 s, demonstrating its robust nonvolatility. The BL voltage differences of both data states, 1 and 0, over a wide range of time durations under 85°C baking are shown in Figure S18. It should be noted that in our experiments, a low V_{dd} of 0.9 V was used for the low-power normal mode, while a higher operation voltage of 1.8 V was required for store and recall operations to ensure sufficient polarization switching in the two FeCaps (Figure S19).[62] These store and recall operation voltages can be reduced by scaling down the HZO thickness[63] or reducing the coercive field (V_c).[64] Maintaining V_{dd} below the V_c (1.2 V in this work) of FeCaps during read and write operations is preferable so that the undesired ferroelectric switching is suppressed and the endurance of FeCaps can be enhanced.[32] The endurance test of our W/HZO/W FeCap is shown in Figure S20, and through further process optimizations, robust endurance beyond 10^{12} cycles[41] can be expected. Our time-dependent dielectric breakdown (TDDB) measurement[65] indicates that reliability of over 10 years can be guaranteed to operate at a V_{dd} of 0.9 V for the normal SRAM mode (Figure S21).[62]

The static noise margin refers to the minimum level of noise that SRAM can withstand without changing its logic state.[66] Figures 4G–4I represent the measured HNM, WNM, and RNM of our M3D nvSRAM, respectively. Even under a low V_{dd} of 0.9 V, reliable SRAM operations characterized by sufficiently high noise margins (HNM of 0.28 V, WNM of 0.45 V, and RNM of 0.26 V) can be achieved. These excellent noise margins are attributed to the precise modulation of the IGZO TFT V_{th} and the optimized FeCap characteristics, which collaboratively enhance the overall data stability and device reliability. As WNM and HNM typically have larger values, enhancing RNM is more essential for maintaining optimal SRAM stability.[21] The RNM is further evaluated with various ratios of channel widths between pass and driver TFTs (W_{Pass}/W_{Driver}), gate voltages of pass TFTs (WL), and V_{dd}s, as depicted in Figures 4J–4L. Although increasing the W_{Pass}/W_{Driver} ratios and WL voltage can boost write/read speeds, these changes reduce the RNM because the Q/Qbar node voltages become more unstable by interference from the BL/BLbar voltages through pass TFTs. It is notable that the RNM of our 4T2C nvSRAM reaches an optimized value at V_{dd} of ~1.1 V and then decreases, which will be discussed in the following section.

Performance evaluation and design space exploration of M3D-4T2C nvSRAM

We have developed device models of HZO-based FeCaps and IGZO TFTs (see methods for details) for comprehensive device-level simulations to fully explore the potential of our 4T2C nvSRAM. Relying on these models, we investigated challenges associated with device scaling and the parameter matching between these components. Then, strategies were explored to optimize memory cell dimensions while preserving high performance. These efforts are essential for 4T2C nvSRAM technology to meet critical requirements of high-density, high-speed, and low-power applications.

The design space of normal SRAM operations is first explored. The read and write operations are conducted in the same manner as those in conventional SRAMs. Intuitively, integrating

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

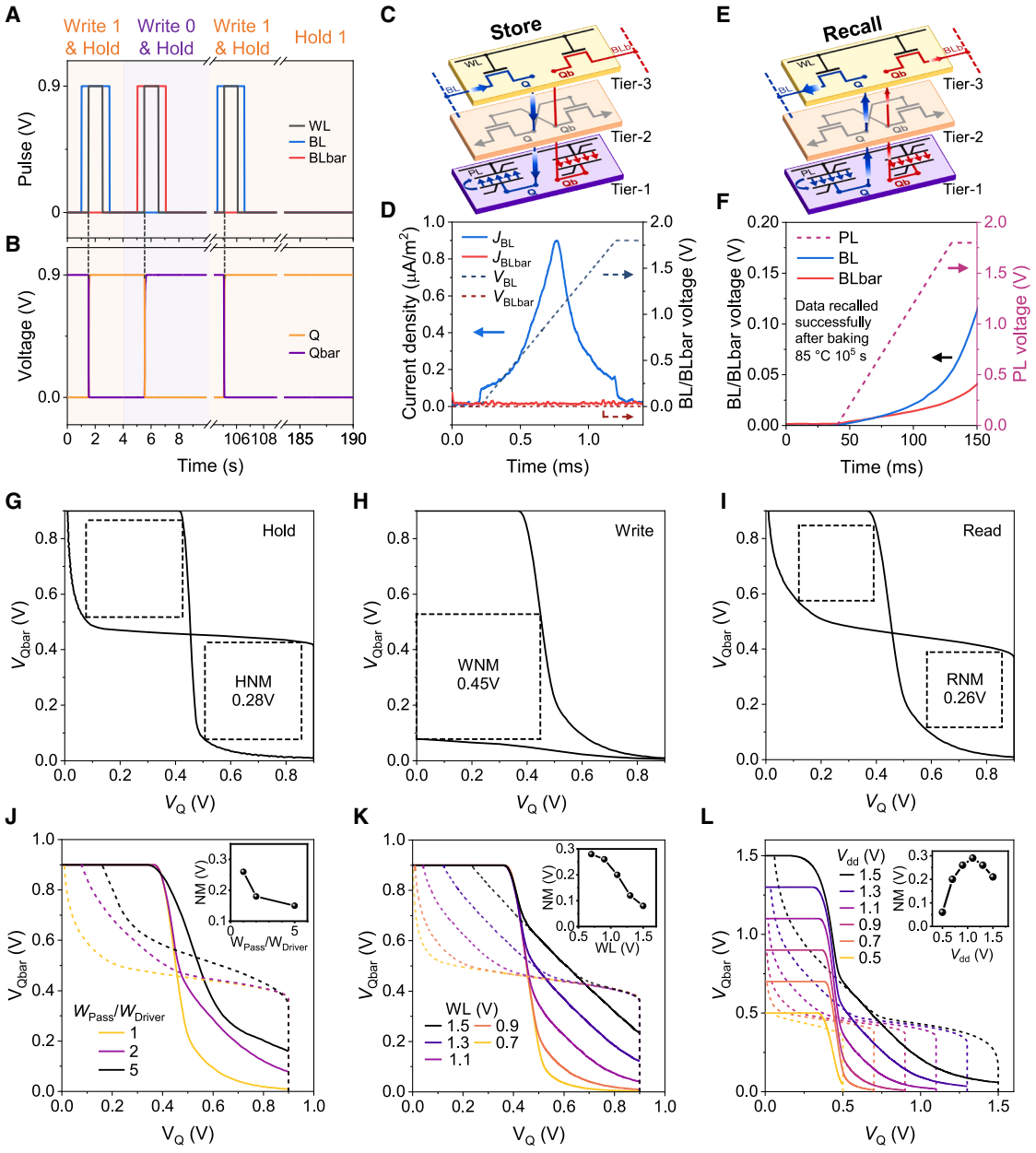


Figure 4. Functionality and reliability demonstration of M3D-4T2C nvSRAM

(A and B) Applied waveforms of WL, BL, and BLbar (A) with measured Q and Qbar node voltages (B) during normal SRAM operations. Both write 1 and 0 operations are demonstrated, and the states can be stably held over 100 s.

(C and D) The schematic of the store operation of data 1 (C) with measured current density at BL/BLbar when WL is turned on and PL is pulled down to 0 V (D). The two FeCaps are switched to different polarization states as Q/Qbar is at high/low voltage level, respectively.

(E and F) The schematic of the recall operation by pulling PL to high and turning on WL (E) with measured BL/BLbar voltage during the recall of data 1 after the device baked at 85°C for 10^5 s (F). The opposite polarization states of FeCaps lead to a sensible voltage difference between BL and BLbar through capacitance coupling.

(G–I) Measured butterfly curves for static noise margins at V_{dd} = 0.9 V with HNM = 0.28 V (G), WNM = 0.45 V (H), and RNM = 0.26 V (I).

(J–L) Measured RNMs with varied channel-width ratios between pass and driver TFTs (W_{Pass}/W_{Driver}) (W_{Driver} kept at 4 μm and V_{dd} = 0.9 V) (J), gate voltages of pass TFTs (V_{dd} = 0.9 V) (K), and V_{dd}s (L). Extracted RNM values are shown as the insets.

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

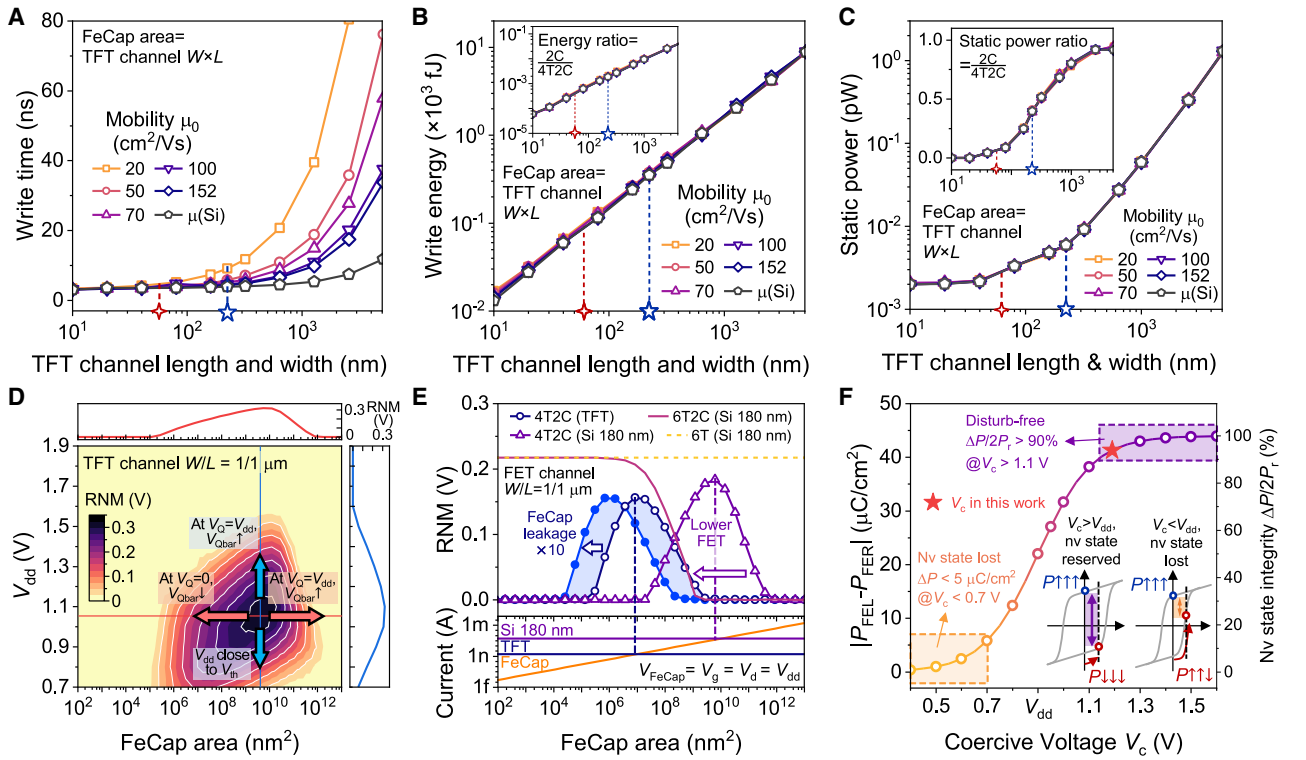


Figure 5. Proposed design space for M3D-4T2C nvSRAM

(A–C) Design space of write time (A), write energy (B), and static power consumption (C) of 4T2C nvSRAM cell with various FeCap/TFT device sizes and IGZO mobilities (μ_0). FeCap and TFT channel areas are scaled simultaneously. Dual V_{dd}s are used during the write operations in (A) and (B), with 1.8 V for WL and 0.9 V for PL/BL/BLbar. Single V_{dd} of 0.9 V is set during the hold operation in (C). Red stars (blue pentagrams) indicate the smallest TFT channel (FeCap) area reported so far.[13,14] Insets of (B) and (C): energy and static power ratio of FeCap/entire 4T2C device. Sub-10 ns write speed can be achieved with both FeCap and TFT channel areas $<10^5$ nm^2 or mobility >100 cm^2/Vs.

(D) RNM variation with different FeCap areas and V_{dd}.

(E) RNM dependence on FeCap area for different nvSRAM and SRAM technologies. RNM window shifts to the left direction (smaller FeCap area) with increased FeCap leakage (blue arrow in the top image) and the use of IGZO TFT (vs. Si-based FET, purple arrow in the top image). Optimized RNM is achieved when FeCap and FET currents ($V_{FeCap} = V_g = V_d = V_{dd}$) are comparable.

(F) Integrity of nonvolatile states as a function of V_c during normal SRAM operations at V_{dd} of 0.9 V, characterized by the difference of FeCap polarization: $\Delta P = |P_{FEL} - P_{FER}|$. FeCap states are reserved when $V_c > V_{dd}$ (left schematic in the inset) and lost when $V_c < V_{dd}$ (right schematic in the inset).

FeCaps into 4T2C devices does not significantly impact the read operation, as the voltages at the data nodes Q and Qbar, which are directly connected to the FeCaps, remain stable throughout the process (Figure S3B). It should be noted that this behavior contrasts with that of 6T2C nvSRAMs, where the integrated FeCaps offer additional paths for discharging the BLs, enhancing the read speed.[32,67] We then primarily focus on analyzing the write operation. The write speed and associated energy consumption are simulated as functions of IGZO mobility and device area, considering both simultaneous scaling of FeCap and TFT channel area (Figures 5A and 5B) and solely scaling FeCap (Figures S4A and S4B). As a result, enhancing IGZO mobility or reducing the device area can coherently boost the write speed and reduce energy consumption. The integration of FeCaps introduces two capacitances directly connected to the storage nodes Q and Qbar. The charging/discharging of linear capacitances of FeCaps slows down the voltage flipping of Q and Qbar nodes, resulting in a write latency (see simulated node voltage evolution and FeCap switching in Figure S22). Therefore, this side effect can be reduced by FeCap scaling, which enhances the write speed and lowers the energy consumption (Figures S4A and S4B). When both FeCap and TFT channel areas are sufficiently small, a sub-10 ns write speed can be achieved (Figure 5A), and the TFTs become the primary factor in energy cost (inset of Figure 5B). When the FeCap area is smaller than the TFT channel, the write speed and energy consumption are influenced predominantly by the TFT channel area (Figures S4A and S4B). Therefore, we can expect that at the advanced technology nodes, the degradation in write operation can be effectively mitigated.

Figures 5C and S4C show the simulated static power of 4T2C cells with various device sizes and mobilities. Attributed to the extremely low off-leakage current in IGZO-TFTs and ultra-low leakage from FeCaps, the static power consumption is several orders of magnitude lower than a 180 nm Si-based 6T SRAM.[32] Reducing the FeCap area results in a suppression of their leakage current, which in turn leads to gradually lower power consumption compared to TFTs (insets of Figures 5C and S4C). We

Please cite this article in press as: Liu et al., High-density, nonvolatile SRAM using monolithic 3D integration of InGaZnO thin-film transistors and $Hf_{0.5}Zr_{0.5}O_2$-based ferroelectric capacitors, Device (2025), https://doi.org/10.1016/j.device.2025.100833

Table 1. Summary of optimization strategies toward high-performance 4T2C nvSRAMs

	FeCap area	V_{dd}	μ_0
Operation speed	↓	↑	↑
Energy efficiency	↓	↓	↑
Static power	↓	↓	–
Noise margin	↗↘	↗↘	–

↑/↓, positive/negative correlation; ↗↘, non-monotonic, optimal point exists; –, limited effect.

observe that based on the most advanced technology reported so far, including the smallest TFT channel area (width/length = 80/40 nm, red stars in Figures 5A–5C),[13] the smallest FeCap area (5×10^4 nm^2, blue pentagrams in Figures 5A–5C),[14] and extremely high mobility (μ_0 = 152 cm^2/Vs),[68] accompanied by our FeCap thickness of 7 nm and sufficiently low V_{dd} of 0.9 V, we can expect a write speed of 3.7 ns, write energy of 103 fJ, static power of 3.2 fW, and recall + store energy of 230 fJ. In particular, if the mobility can approach that of Si, then both the write time and energy can be comparable to those of a normal Si-based 6T SRAM, indicating that the M3D 2T-footprint 4T2C can be a competitive alternative to standard SRAMs.

The noise margin, as discussed in the previous section, has been treated as a critical metric for data stability. An optimal RNM that balances the FeCap area and V_{dd} can be identified (Figure 5D), consistent with our measured RNM shown in Figure 4L. More specifically, on the one hand, the FeCap area should be tailored because its effective resistance affects the voltage division between FeCaps and TFTs. Purely reducing the FeCap area degrades the RNM by lowering the V_{Qbar} at V_Q = 0 V, whereas increasing the FeCap area raises the V_{Qbar} at $V_Q = V_{dd}$. On the other hand, adjusting V_{dd} modulates the current drive capability of pass TFTs. Elevating V_{dd} effectively enhances the device reliability by booting the on current (I_{on}) of TFTs but may undermine the RNM as V_{Qbar} rises at $V_Q = V_{dd}$. We also investigated optimization strategies for ensuring robust operations regarding the couplings of V_{dd}s and device parameters (Note S4; Figures S5 and S6). Figure 5E compares the RNMs of various types of SRAMs, including 4T2C nvSRAM using TFTs, as well as 4T2C nvSRAM, 6T2C nvSRAM, and traditional 6T SRAM using 180 nm Si-based FETs. The maximum RNM in 4T2C nvSRAMs is smaller than that in 6T-based SRAMs due to the highly non-linear *I-V* characteristics of p-type load MOSFETs compared to that of the resistor-load-like property of FeCaps.[66] In 4T2C nvSRAMs, an optimal RNM can be achieved when the currents flowing through FeCaps and TFTs are comparable (bottom of Figure 5E). Since the reduction of the FeCap area is critical for improving the speed and energy consumption of our nvSRAM, as discussed earlier, the required FeCap area can be minimized by enhancing the FeCap leakage current (blue arrow in the top image of Figure 5E), attaching resistors in parallel with FeCaps, or suppressing the FET current (purple arrow in the top image of Figure 5E, comparing the use of Si-based FETs and IGZO TFTs).

Typically, to prevent unexpected data loss, a dedicated circuit design is responsible for automatically executing a store operation at power failure. Here, we propose that introducing a store operation after each recall or write can preserve the stored nonvolatile states during the subsequent normal SRAM operations, provided that $V_{dd} < V_c$ is maintained (Figure 5F). Otherwise, only a minor portion of the ferroelectric polarization can be retained in the same direction, causing the loss of the nonvolatile states (right schematic as the inset in Figure 5F). Then, adopting a larger-V_c FeCap allows for a greater P_r to be retained during normal SRAM operations, and in this work, a V_c of 1.2 V is sufficiently high to suppress undesired polarization loss, thereby improving device endurance (left schematic as the inset in Figure 5F).

The trade-offs and optimization strategies are concisely outlined in Figure S23 and Table 1, focusing on key performance metrics, including operation speed, energy consumption, and data stability. The systematic evaluation of the overall reliability of the 4T2C nvSRAM, based on the performance of both TFTs and FeCaps, is illustrated in Figure S5. Achieving optimal performance in the 4T2C nvSRAM requires enhancing IGZO mobility while cautiously managing device area, leakage current, and supply voltage. Our research into parameter optimization ensures seamless coordination between memory and logic components, thereby enabling higher integration density and enhanced device efficiency for the future mass production of M3D-4T2C nvSRAM technology.

Conclusions and outlook

We presented a 2T-footprint M3D-integrated 4T2C nvSRAM through 3-tier stacking of optimized IGZO TFTs and HZO-based FeCaps with BEOL-compatible temperatures (≤400°C). Our design addresses the limitations of conventional Si-based SRAMs by achieving nonvolatile data retention after exposure to 85°C for 10^5 s and reducing the cell size through vertical stacking and minimizing transistor numbers. In addition to the functionality demonstration, an HNM of 0.28 V, a WNM of 0.45 V, and an RNM of 0.26 V are achieved with a low V_{dd} of 0.9 V. Based on comprehensive simulations, we have provided detailed design space for the 4T2C nvSRAM toward higher operation speed, lower energy consumption, and more robust data stability. Table S1 benchmarks our results against state-of-the-art experimentally demonstrated nvSRAMs in the literature.[23–26,30] Our work supports the M3D-4T2C nvSRAM as a compelling memory solution for 3D-stacked emerging computing systems, requiring high density, fast speed, superior write/read endurance, robust nonvolatility, and low power consumption, but also underscores the potential of M3D integration with emerging logic and memory devices for more innovative device technologies.

Future studies may advance this work in two main directions, including device innovation and process optimization, aiming for full compatibility with advanced CMOS technology nodes while achieving continuous improvements in density, speed, and power consumption. On the one hand, adopting OS transistors with superior mobility and stability can enable faster and more energy-efficient operations, and further scaling of FeCaps with lower coercive voltages and enhanced reliability may elevate nonvolatile performance. On the other hand, breakthroughs in M3D integration, particularly via vertical channel transistors and 3D FeCaps, are poised to push integration limits further, paving the way for highly compact and efficient system architectures.

METHODS

M3D fabrication of 4T2C nvSRAM

BEOL-compatible M3D fabrication processes for 4T2C nvSRAMs were developed. Further details are provided in Note S1 and Figure S1.

Characterization techniques

TEM and EDS were employed for microscopic imaging, GIXRD was used to characterize crystal structures, D-SIMS was adopted to analyze hydrogen depth profiles, and electrical measurements were conducted to evaluate electrical characteristics of FeCaps, TFTs, and operations of 4T2C devices. Further information on the characterization techniques is available in Note S5.

Device modeling and simulations

Models for HZO-based FeCaps and IGZO TFTs were developed. The FeCap module captures polarization switching and leakage current behaviors, and the IGZO TFT module reproduces *I-V* and *C-V* characteristics. Detailed modeling formulations are provided in Note S6 and Figures S7 and S8.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hao Jiang (haoj@fudan.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data generated during the study are available in the paper and supplemental information.

ACKNOWLEDGMENTS

This work was supported by part of the National Key Research and Development Program of China (grant 2022YFB3608400), the National Natural Science Foundation of China (grants 62404048, 6240030380, and 62404053), the China Postdoctoral Science Foundation (grant 2023M730620), and the Ministry of Education (MOE) innovation platform.

AUTHOR CONTRIBUTIONS

Qihan Liu performed the experiments. Y.L. and L.L. developed the model and carried out the simulation work. H.Z. and Z.C. contributed to the circuit design. Y.Y. analyzed the TEM images. H.J., Y.W., and Qi Liu supervised the research. Qihan Liu, Y.L., H.J., M.S., C.L., P.L., Y.W., and Qi Liu wrote and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.device.2025.100833.

Received: January 27, 2025
Revised: January 31, 2025
Accepted: May 16, 2025

REFERENCES

1. Yuvaraja, S., Faber, H., Kumar, M., Xiao, N., Maciel García, G.I., Tang, X., Anthopoulos, T.D., and Li, X. (2024). Three-dimensional integrated metal-oxide transistors. Nat. Electron. *7*, 768–776. https://doi.org/10.1038/s41928-024-01205-0.
2. Pendurthi, R., Sakib, N.U., Sadaf, M.U.K., Zhang, Z., Sun, Y., Chen, C., Jayachandran, D., Oberoi, A., Ghosh, S., Kumari, S., et al. (2024). Monolithic three-dimensional integration of complementary two-dimensional field-effect transistors. Nat. Nanotechnol. *19*, 970–977. https://doi.org/10.1038/s41565-024-01705-2.
3. Yang, H., Li, Y., Tang, J., An, R., Zhang, Y., Gao, L., Gao, N., Xu, H., Du, Y., and Liu, Z. (2024). Monolithic 3D Integration of Analog RRAM-Based Fully Weight Stationary and Novel CFET 2T0C-Based Partially Weight Stationary for Accelerating Transformer. In 2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.
4. Shen, Y., Zhu, K., Xiao, Y., Waldhör, D., Basher, A.H., Knobloch, T., Pazos, S., Liang, X., Zheng, W., Yuan, Y., et al. (2024). Two-dimensional-materials-based transistors using hexagonal boron nitride dielectrics and metal gate electrodes with high cohesive energy. Nat. Electron. *7*, 856–867. https://doi.org/10.1038/s41928-024-01233-w.
5. Jayachandran, D., Pendurthi, R., Sadaf, M.U.K., Sakib, N.U., Pannone, A., Chen, C., Han, Y., Trainor, N., Kumari, S., Mc Knight, T.V., et al. (2024). Three-dimensional integration of two-dimensional field-effect transistors. Nature *625*, 276–281. https://doi.org/10.1038/s41586-023-06860-5.
6. Yang, J.-E., Jang, Y., Han, N.-R., Sung, H.-J., Kim, J.-k., Cha, Y., Lee, K.-H., Jung, K., Jung, M., and Lee, W. (2024). A-IGZO FETs with High Current and Remarkable Stability for Vertical Channel Transistor (VCT)/3D DRAM Applications. In 2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.
7. Lee, F.-M., Tseng, P.-H., Lin, Y.-Y., Lin, Y.-H., Weng, W.-L., Lin, N.-C., Sung, P.-J., Wu, C.-T., Yang, C.-C., and Wu, W.-F. (2024). Bit-Cost-Scalable 3D DRAM Architecture and Unit Cell First Demonstrated with Integrated Gate-Around and Channel-Around IGZO FETs. In 2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.
8. Kim, I.-J., Kim, M.-K., and Lee, J.-S. (2023). Highly-scaled and fully-integrated 3-dimensional ferroelectric transistor array for hardware implementation of neural networks. Nat. Commun. *14*, 504. https://doi.org/10.1038/s41467-023-36270-0.
9. Du, Y., Tang, J., Li, Y., Xi, Y., Li, Y., Li, J., Huang, H., Qin, Q., Zhang, Q., Gao, B., et al. (2024). Monolithic 3D Integration of Analog RRAM-Based Computing-in-Memory and Sensor for Energy-Efficient Near-Sensor Computing. Adv. Mater. *36*, e2302658. https://doi.org/10.1002/adma.202302658.
10. Dutta, S., Ye, H., Chakraborty, W., Luo, Y.C., Jose, M.S., Grisafe, B., Khanna, A., Lightcap, I., Shinde, S., Yu, S., et al. (2020). Monolithic 3D Integration of High Endurance Multi-Bit Ferroelectric FET for Accelerating Compute-In-Memory. In 2020 International Electron Devices Meeting (IEDM), pp. 36.34.31–36.34.34.
11. Gu, C., Hu, Q., Zhu, S., Li, Q., Zeng, M., Kang, J., Tong, A., and Wu, Y. (2024). First Experimental Demonstration of 3D-Stacked 2T0C DRAM Cells Based on Indium Tin Oxide Channel. IEEE Electron Device Lett. *45*, 1764–1767. https://doi.org/10.1109/led.2024.3443512.
12. Choi, S., Shin, J., Park, G., Eo, J.S., Jang, J., Yang, J.J., and Wang, G. (2024). 3D-integrated multilayered physical reservoir array for learning and forecasting time-series information. Nat. Commun. *15*, 2044. https://doi.org/10.1038/s41467-024-46323-7.
13. Subhechha, S., Rassoul, N., Belmonte, A., Hody, H., Dekkers, H., van Setten, M.J., Chasin, A., Sharifi, S.H., Sutar, S., Magnarin, L., et al. (2022). Ultra-low Leakage IGZO-TFTs with Raised Source/Drain for $V_{th} > 0$ V and $I_{on} > 30$ μA/μm. In 2022 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 292–293.
14. Ramaswamy, N., Calderoni, A., Zahurak, J., Servalli, G., Chavan, A., Chhajed, S., Balakrishnan, M., Fischer, M., Hollander, M., and Ettisserry,

D. (2023). NVDRAM: A 32Gb dual layer 3D stacked non-volatile ferroelectric memory with near-DRAM performance for demanding AI workloads. In 2023 International Electron Devices Meeting (IEDM), pp. 1–4.

15. Lin, P., Li, C., Wang, Z., Li, Y., Jiang, H., Song, W., Rao, M., Zhuo, Y., Upadhyay, N.K., Barnell, M., et al. (2020). Three-dimensional memristor circuits as complex neural networks. Nat. Electron. *3*, 225–232. https://doi.org/10.1038/s41928-020-0397-9.
16. Jayachandran, D., Sakib, N.U., and Das, S. (2024). 3D integration of 2D electronics. Nat. Rev. Electr. Eng. *1*, 300–316. https://doi.org/10.1038/s44287-024-00038-5.
17. Lu, A., Lee, J., Kim, T.-H., Karim, M.A.U., Park, R.S., Simka, H., and Yu, S. (2024). High-speed emerging memories for AI hardware accelerators. Nat. Rev. Electr. Eng. *1*, 24–34. https://doi.org/10.1038/s44287-023-00002-9.
18. Yu, D. (2021). TSMC packaging technologies for chiplets and 3D. In Proceedings of the IEEE Hot Chips, pp. 47–56.
19. Radosavljevic, M., and Kavalieros, J. (2022). 3D-stacked CMOS takes Moore's law to new heights. In IEEE. Spectrum. https://spectrum.ieee.org/3d-cmos.
20. Cao, W., Bu, H., Vinet, M., Cao, M., Takagi, S., Hwang, S., Ghani, T., and Banerjee, K. (2023). The future transistors. Nature *620*, 501–515. https://doi.org/10.1038/s41586-023-06576-6.
21. Lu, Y.-C., Huang, J.-K., Chao, K.-Y., Li, L.-J., and Hu, V.P.-H. (2024). Projected performance of Si-and 2D-material-based SRAM circuits ranging from 16 nm to 1 nm technology nodes. Nat. Nanotechnol. *19*, 1066–1072. https://doi.org/10.1038/s41565-024-01693-3.
22. Jhang, C.-J., Xue, C.-X., Hung, J.-M., Chang, F.-C., and Chang, M.-F. (2021). Challenges and Trends of SRAM-Based Computing-In-Memory for AI Edge Devices. IEEE Trans. Circ. Syst. I. *68*, 1773–1786. https://doi.org/10.1109/tcsi.2021.3064189.
23. Kobayashi, M., Ueyama, N., and Hiramoto, T. (2017). A nonvolatile SRAM integrated with ferroelectric HfO_2 capacitor for normally-off and ultralow power IoT application. In 2017 Symposium on VLSI Technology, pp. T156–T157.
24. Hsu, M.-Y., Liao, C.-F., Shih, Y.-H., Lin, C.J., and King, Y.-C. (2017). A RRAM Integrated 4T SRAM with Self-Inhibit Resistive Switching Load by Pure CMOS Logic Process. Nanoscale Res. Lett. *12*, 418. https://doi.org/10.1186/s11671-017-2191-9.
25. Lee, A., Chang, M.-F., Lin, C.-C., Chen, C.-F., Ho, M.-S., Kuo, C.-C., Tseng, P.-L., Sheu, S.-S., and Ku, T.-K. (2015). RRAM-based 7T1R nonvolatile SRAM with 2x reduction in store energy and 94x reduction in restore energy for frequent-off instant-on applications. In 2015 Symposium on VLSI Circuits (VLSI Circuits), pp. C76–C77.
26. Ohsawa, T., Koike, H., Miura, S., Honjo, H., Kinoshita, K., Ikeda, S., Hanyu, T., Ohno, H., and Endoh, T. (2013). A 1 Mb Nonvolatile Embedded Memory Using 4T2MTJ Cell With 32 b Fine-Grained Power Gating Scheme. IEEE J. Solid State Circ. *48*, 1511–1520. https://doi.org/10.1109/jssc.2013.2253412.
27. You, W.-X., Su, P., and Hu, C. (2020). A New 8T Hybrid Nonvolatile SRAM With Ferroelectric FET. IEEE J. Electron Devices Soc. *8*, 171–175. https://doi.org/10.1109/jeds.2020.2972319.
28. Huang, K., Ha, Y., Zhao, R., Kumar, A., and Lian, Y. (2014). A Low Active Leakage and High Reliability Phase Change Memory (PCM) Based Non-Volatile FPGA Storage Element. IEEE Trans. Circ. Syst. I. *61*, 2605–2613. https://doi.org/10.1109/tcsi.2014.2312499.
29. Nouri, S., and Iyer, S.S. (2023). An 8T eNVSRAM Macro in 22nm FDSOI Standard Logic with Simultaneous Full-Array Data Restore for Secure IoT Devices. In 2023 IEEE International Solid-State Circuits Conference (ISSCC), pp. 434–436.
30. Shuto, Y., Okuno, J., Yonai, T., Ono, R., Reinig, P., Lederer, M., Seidel, K., Alcala, R., Mikolajick, T., and Schroeder, U. (2024). HZO-based Nonvolatile SRAM Array with 100% Bit Recall Yield and Sufficient Retention Time at 85 °C. In 2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.
31. Muller, J., Boscke, T.S., Schroder, U., Mueller, S., Brauhaus, D., Bottger, U., Frey, L., and Mikolajick, T. (2012). Ferroelectricity in simple binary ZrO_2 and HfO_2. Nano Lett. *12*, 4318–4323. https://doi.org/10.1021/nl302049k.
32. Li, Y., Zhou, K., Jiang, H., Huang, Z., Zhang, X., Wei, Y., and Liu, Q. (2024). Working Principles and Performance Optimization of Nonvolatile 6T2C-SRAM With Hafnia-Based Ferroelectric Capacitors. IEEE Trans. Electron. Dev. *71*, 4631–4636. https://doi.org/10.1109/ted.2024.3408778.
33. Jiang, H., Li, O., Chen, W., and Ma, T.P. (2020). Dual-Storage-Port Nonvolatile SRAM Based on Back-End-of-the-Line Processed $Hf_{0.5}Zr_{0.5}O_2$ Ferroelectric Capacitors Towards 3D Selector-Free Cross-Point Memory. IEEE J. Electron Devices Soc. *8*, 935–938. https://doi.org/10.1109/jeds.2020.3019024.
34. Charnas, A., Zhang, Z., Lin, Z., Zheng, D., Zhang, J., Si, M., and Ye, P.D. (2024). Extremely Thin Amorphous Indium Oxide Transistors. Adv. Mater. *36*, e2304044. https://doi.org/10.1002/adma.202304044.
35. Geng, D., Wang, K., Li, L., Myny, K., Nathan, A., Jang, J., Kuo, Y., and Liu, M. (2023). Thin-film transistors for large-area electronics. Nat. Electron. *6*, 963–972. https://doi.org/10.1038/s41928-023-01095-8.
36. Hosono, H. (2018). How we made the IGZO transistor. Nat. Electron. *1*, 428. https://doi.org/10.1038/s41928-018-0106-0.
37. Liu, S.-E., Yu, M.-J., Lin, C.-Y., Ho, G.-T., Cheng, C.-C., Lai, C.-M., Lin, C.-J., King, Y.-C., and Yeh, Y.-H. (2011). Influence of Passivation Layers on Characteristics of a-InGaZnO Thin-Film Transistors. IEEE Electron Device Lett. *32*, 161–163. https://doi.org/10.1109/led.2010.2091620.
38. Nguyen, T.T.T., Aventurier, B., Terlier, T., Barnes, J.-P., and Templier, F. (2015). Impact of passivation conditions on characteristics of bottom-gate IGZO thin-film transistors. J. Display Technol. *11*, 554–558. https://doi.org/10.1109/JDT.2015.2396476.
39. Aman, S.G.M., Koretomo, D., Magari, Y., and Furuta, M. (2018). Influence of Deposition Temperature and Source Gas in PE-CVD for SiO_2 Passivation on Performance and Reliability of In–Ga–Zn–O Thin-Film Transistors. IEEE Trans. Electron. Dev. *65*, 3257–3263. https://doi.org/10.1109/ted.2018.2841978.
40. Lee, Y., Goh, Y., Hwang, J., Das, D., and Jeon, S. (2021). The Influence of Top and Bottom Metal Electrodes on Ferroelectricity of Hafnia. IEEE Trans. Electron. Dev. *68*, 523–528. https://doi.org/10.1109/ted.2020.3046173.
41. Chen, G.-H., Liao, Y.-T., Zhao, Z., Chen, Y.-R., Chen, Y.-W., Chen, W.-J., Hsu, W.-T., Lu, H.-Y., Liu, M.-C., Chen, Y.-A., et al. (2024). Uniform and Fatigue-Free Ferroelectric HZO with Record EBD of 6.3MV/cm and Record Final 2Pr of 64uC/cm^2 at Record 5E12 Endurance Using Low Lattice Misfit (2.9%) β−W. In 2024 IEEE International Electron Devices Meeting (IEDM), pp. 1–4.
42. Park, M.H., Kim, H.J., Kim, Y.J., Moon, T., Kim, K.D., Lee, Y.H., Hyun, S.D., and Hwang, C.S. (2015). Study on the internal field and conduction mechanism of atomic layer deposited ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ thin films. J. Mater. Chem. C *3*, 6291–6300. https://doi.org/10.1039/C5TC01074H.
43. Kong, Q., Liu, L., Han, K., Sun, C., Jiao, L., Zhou, Z., Zheng, Z., Liu, G., Xu, H., Zhang, J., et al. (2024). Discovering the Impact of Cooling Scheme During Annealing: A New Knob for Achieving Thermally Stable IGZO FETs. IEEE Trans. Electron. Dev. *71*, 5425–5431. https://doi.org/10.1109/ted.2024.3433832.
44. Wahid, S., Daus, A., Kwon, J., Qin, S., Ko, J.-S., Wong, H.S.P., and Pop, E. (2023). Effect of Top-Gate Dielectric Deposition on the Performance of Indium Tin Oxide Transistors. IEEE Electron Device Lett. *44*, 951–954. https://doi.org/10.1109/led.2023.3265316.
45. Chowdhury, M.D.H., Mativenga, M., Jae Gwang, U., Mruthyunjaya, R.K., Heiler, G.N., Tredwell, T.J., and Jin, J. (2015). Effect of SiO_2 and SiO_2/SiN_x Passivation on the Stability of Amorphous Indium-Gallium Zinc-Oxide Thin-Film Transistors Under High Humidity. IEEE Trans. Electron. Dev. *62*, 869–874. https://doi.org/10.1109/ted.2015.2392763.
46. Nomura, K., Kamiya, T., and Hosono, H. (2012). Stability and high-frequency operation of amorphous In–Ga–Zn–O thin-film transistors with

various passivation layers. Thin Solid Films *520*, 3778–3782. https://doi.org/10.1016/j.tsf.2011.10.068.

47. Lee, D.Y., Park, J., Lee, S., Myoung, S.J., Lee, H., Bae, J.-H., Choi, S.-J., Kim, D.M., Kim, C., and Kim, D.H. (2024). Influence of RF power in the sputter deposition of amorphous InGaZnO film on the transient drain current of amorphous InGaZnO thin-film transistors. Solid State Electron. *216*, 108921. https://doi.org/10.1016/j.sse.2024.108921.

48. Xu, W., Xu, M., Jiang, J., Xu, S., and Feng, X. (2019). Impact of Sputtering Power on Amorphous In–Al–Zn–O Films and Thin Film Transistors Prepared by RF Magnetron Sputtering. IEEE Trans. Electron. Dev. *66*, 2219–2223. https://doi.org/10.1109/ted.2019.2906892.

49. Si, M., Hu, Y., Lin, Z., Sun, X., Charnas, A., Zheng, D., Lyu, X., Wang, H., Cho, K., and Ye, P.D. (2021). Why In-$_2O_3$ Can Make 0.7 nm Atomic Layer Thin Transistors. Nano Lett. *21*, 500–506. https://doi.org/10.1021/acs.nanolett.0c03967.

50. Samanta, S., Han, K., Sun, C., Wang, C., Kumar, A., Thean, A.V.-Y., and Gong, X. (2021). Amorphous InGaZnO Thin-Film Transistors With Sub-10-nm Channel Thickness and Ultrascaled Channel Length. IEEE Trans. Electron. Dev. *68*, 1050–1056. https://doi.org/10.1109/ted.2020.3048920.

51. Lee, T., and Oh, S. (2024). Improvement of Amorphous InGaZnO Thin-Film Transistor Reliability and Electrical Performance Using ALD SiO_2 Interfacial Layer on PECVD SiO_2 Gate Insulator. IEEE Trans. Electron. Dev. *71*, 1926–1931. https://doi.org/10.1109/ted.2024.3355025.

52. Jeon, J.K., Um, J.G., Lee, S., and Jang, J. (2017). Control of O-H bonds at a-IGZO/SiO_2 interface by long time thermal annealing for highly stable oxide TFT. AIP Adv. *7*. https://doi.org/10.1063/1.5008435.

53. Seshadri, A., Blake, T.G., and Eliason, J.R. (2005). NON-VOLATILE SRAM. United States Patent USOO6980459B2.

54. Wu, Q., Cao, Y., Luo, Q., Jiang, H., Han, Z., Han, Y., Dou, C., Lv, H., Liu, Q., Yang, J., and Liu, M. (2024). 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. IEEE J. Solid State Circ. *59*, 208–218. https://doi.org/10.1109/jssc.2023.3320659.

55. Kim, J., Han, M., Bang, J., Lim, Y., and Choi, J. (2025). A Command-Aware Hybrid LDO for Advanced HBM Interfaces with 150μA Quiescent Current and 20pF On-Chip Capacitor Achieving Sub-10mV Voltage Droop in 400ps Settling Time. In 2025 IEEE International Solid-State Circuits Conference (ISSCC), pp. 1–3.

56. Okuno, J., Kunihiro, T., Konishi, K., Maemura, H., Shuto, Y., Sugaya, F., Materano, M., Ali, T., Kuehnel, K., and Seidel, K. (2020). SoC compatible 1T1C FeRAM memory array based on ferroelectric $Hf_{0.5}Zr_{0.5}O_2$. In 2020 IEEE Symposium on VLSI Technology, pp. 1–2.

57. Hwang, S., Jang, H., Lee, K., Jung, L., Yoon, J., Kwon, J.-D., Song, K., Kim, Y., and Hwang, H. (2024). Record Endurance (> 10^{12} Cycles), High Polarization (2Pr > 50 μC/cm^2), and 10-Year Data Retention (85°C) in HZO Capacitors with Well-Ordered Ferroelectric Domain Structures via 2D-WS_2 Interface. In 2024 IEEE International Electron Devices Meeting (IEDM), pp. 1–4.

58. Walke, A.M., Popovici, M.I., Sharifi, S.H., Demir, E.C., Puliyalil, H., Bizindavyi, J., Yasin, F., Clima, S., Fantini, A., Belmonte, A., et al. (2024). La Doped HZO-Based 3D-Trench Metal-Ferroelectric-Metal Capacitors With High-Endurance (>10^{12}) for FeRAM Applications. IEEE Electron Device Lett. *45*, 578–581. https://doi.org/10.1109/led.2024.3368225.

59. Yuan, P., Wang, B., Yang, Y., Lv, S., Wang, Y., Xu, Y., Jiang, P., Chen, Y., Dang, Z., Ding, Y., et al. (2022). Enhanced Remnant Polarization (30 μC/cm^2) and Retention of Ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ by NH_3 Plasma Treatment. IEEE Electron Device Lett. *43*, 1045–1048. https://doi.org/10.1109/led.2022.3178867.

60. Laguerre, J., Martin, S., Coignus, J., Carabasse, C., Bocquet, M., Andrieu, F., and Grenouillet, L. (2024). Data Retention Insights from Joint Analysis on BEOL-Integrated HZO-Based Scaled FeCAPs and 16kbit 1T-1C FeRAM Arrays. In 2024 IEEE International Reliability Physics Symposium (IRPS), pp. 1–7.

61. Gong, T., Xu, L., Wei, W., Jiang, P., Yuan, P., Nie, B., Huang, Y., Wang, Y., Yang, Y., and Gao, J. (2023). First Demonstration of a Design Methodology for Highly Reliable Operation at High Temperature on 128kb 1T1C FeRAM Chip. In 2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.

62. Si, M., Lyu, X., Shrestha, P.R., Sun, X., Wang, H., Cheung, K.P., and Ye, P.D. (2019). Ultrafast measurements of polarization switching dynamics on ferroelectric and anti-ferroelectric hafnium zirconium oxide. Appl. Phys. Lett. *115*, 072107. https://doi.org/10.1063/1.5098786.

63. Tahara, K., Toprasertpong, K., Hikosaka, Y., Nakamura, K., Saito, H., Takenaka, M., and Takagi, S. (2021). Strategy toward HZO BEOL-FeRAM with low-voltage operation ($\leq$ 1.2 V), low process temperature, and high endurance by thickness scaling. In 2021 Symposium on VLSI Technology, pp. 1–2.

64. Wang, Y., Tao, L., Guzman, R., Luo, Q., Zhou, W., Yang, Y., Wei, Y., Liu, Y., Jiang, P., Chen, Y., et al. (2023). A stable rhombohedral phase in ferroelectric $Hf(Zr)_{1+x}O_2$ capacitor with ultralow coercive field. Science *381*, 558–563. https://doi.org/10.1126/science.adf6137.

65. Liu, Z., Cai, P., Yu, S., Han, L., Wang, R., Wu, Y., Ren, P., Ji, Z., and Huang, R. (2021). Investigation of Time Dependent Dielectric Breakdown (TDDB) of $Hf_{0.5}Zr_{0.5}O_2$-Based Ferroelectrics Under Both Forward and Reverse Stress Conditions. IEEE J. Electron Devices Soc. *9*, 735–740. https://doi.org/10.1109/jeds.2021.3103182.

66. Seevinck, E., List, F.J., and Lohstroh, J. (1987). Static-noise margin analysis of MOS SRAM cells. IEEE J. Solid State Circ. *22*, 748–754. https://doi.org/10.1109/JSSC.1987.1052809.

67. Miwa, T., Yamada, J., Koike, H., Toyoshima, H., Amanuma, K., Kobayashi, S., Tatsumi, T., Maejima, Y., Hada, H., and Kunio, T. (2001). NV-SRAM: A nonvolatile SRAM with backup ferroelectric capacitors. IEEE J. Solid State Circ. *36*, 522–527. https://doi.org/10.1109/4.910492.

68. Lin, Z., Niu, C., Jang, H., Kim, T., Zhang, Y., Wang, H., Jeong, C., and Ye, D.P. (2024). Enhancement of In_2O_3 Field-Effect Mobility Up To 152 cm^{2}V^{-1}s^{-1} Using HZO-Based Higher-*k* Linear Dielectric. In 2024 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), pp. 1–2.