Paasio et al. *Microsystems & Nanoengineering* (2026)12:313
https://doi.org/10.1038/s41378-026-01429-4

Microsystems & Nanoengineering
www.nature.com/micronano

ARTICLE Open Access

High-precision hyperdimensional computing enabled by in-memory computing using high-polarization ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ capacitors

Ella Paasio[1]✉, Simon Thomann[2], Anika Anu[1], Xinye Li[1], Rikhard Ranta[1], Padma Srivari[1], Safdar Muhammad[3], Soumen Mazumder[3], Jahra Mariam[3], Andrea Padovani[4], Hussam Amrouch[2], Gaurav Thareja[5] and Sayani Majumdar[1]✉

Abstract

We demonstrate high-precision hyperdimensional computing using an in-memory computing (IMC) architecture based on ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ (HZO) capacitors. By exploiting the high polarization charge density of CMOS back-end-compatible HZO, we achieved 32 well-separated and linearly programmable intermediate states in 10-nm-thick capacitors making them suitable as capacitive IMC elements. In recent times, capacitive IMC emerged as a promising energy- and latency-efficient route for data-intensive computing tasks. However, compute-in-memory elements require non-volatile, reproducible, and multi-bit operation. In this work, we show that through optimized device fabrication without vacuum break between oxide and nitride depositions and tailored thermal engineering, the HZO capacitors can exhibit high remanent polarization ($2P_r = 75\ \mu C/cm^2$). Structural studies highlight a high orthorhombic phase fraction and clean HZO/TiN interface. The intermediate polarization states exhibit controllable, linear, and reproducible capacitance modulation via voltage-driven polarization switching, enabling reliable multi-bit device operation and non-destructive readout. Leveraging these 5-bit ferroelectric capacitors, it is possible to store 15-bit numerical values using only three capacitors to implement high-precision capacitive IMC in a hyperdimensional computing task, achieving improved inference accuracy of 92.3% and 2.3x reduced areal footprint compared to binary encoding. These results highlight the importance of advanced materials engineering to achieve high bit-precision and state linearity in ferroelectric capacitors for scalable capacitive in-memory computing.

Introduction

Data-intensive computing, particularly in machine learning and edge computing, has caused an exponential increase in the demand for energy-efficient, high-performance hardware, especially in the field of non-volatile memories (NVM). In-memory computing (IMC) offers a promising path forward in decreasing data shuttling across the memory–processor interface, enabling massively parallel operations and reducing the memory wall effect[1–5]. However, to fully benefit from analog IMC, memory devices need to support multi-bit storage per cell, as higher bit density directly translates to improved computational throughput while minimizing footprint on chip. These analog storage devices must have well-defined, stable, and reproducible memory states[5]. Further requirements such as non-volatility, analog programmability and compatibility with complementary metal-oxide-semiconductor (CMOS) technology can be fulfilled by different device technologies, out of which ferroelectric-based devices have emerged as compelling candidates due to their extreme energy efficiency[2,3,6–8].

Capacitive in-memory computing leverages charge accumulation and redistribution within memory elements to perform arithmetic operations directly in the analog domain, thereby eliminating the energy and latency

Correspondence: Ella Paasio (ella.paasio@tuni.fi) or Sayani Majumdar (sayani.majumdar@tuni.fi)
[1]Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland
[2]School of Computation, Information and Technology, Techical University of Munich, Munich, Germany
Full list of author information is available at the end of the article

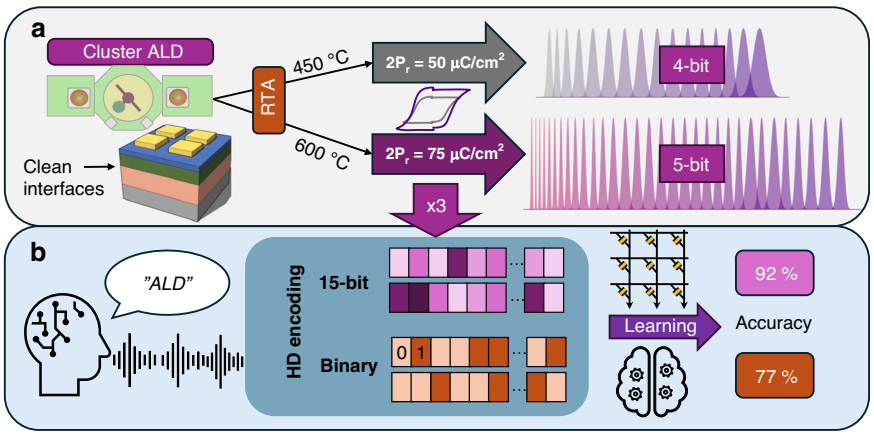

Fig. 1 Overview of fabricated ferroelectric capacitors improving hyperdimensional computing task learning accuracy. **a** The experimental work reported in this study comprises ferroelectric capacitor (FeCAP) device fabrication, structural and electrical characterization, analog state identification and their reliability study. **b** The computational part of the work explores the benefits of using characteristics from the fabricated devices in a hyperdimensional computing scheme

penalties associated with data movement in conventional architectures[9,10]. Unlike resistive approaches that rely on conductance modulation, capacitive schemes exploit voltage-controlled charge storage, offering inherently low leakage and compatibility with high-speed, low-power circuits[3,7,11]. This paradigm is particularly attractive for multiply–accumulate (MAC) operations in neural networks, where weighted sums are realized through controlled charge sharing across capacitive nodes. When implemented using ferroelectric capacitors, the tuneable polarization states provide a natural mechanism for multi-level charge storage, enabling both dense memory integration and analog computation within the same physical device[2,7,12].

Ferroelectric (FE) materials, especially $Hf_{0.5}Zr_{0.5}O_2$ (HZO), compatible with CMOS back-end-of-line (BEOL) integration, have emerged as strong candidates for multibit memory and IMC due to their scalability, fast switching dynamics, and intrinsic non-volatility[6,12–14]. Through compatibility with matured industry-scale atomic layer deposition (ALD), scalability down to a few nanometer thickness[14,15], and the ability to engineer intermediate polarization states through voltage-controlled domain switching[16], the technology is heading towards multi-bit storage within a single cell. Importantly, recent studies indicate that ferroelectric properties in HZO persist over a large temperature window, from 400 K down to deep cryogenic temperature of 4 K with improved endurance and retention properties at low temperature, opening opportunities for integration with cryogenic electronics used in quantum computing, space technologies and high-performance data centers[14,17–19]. However, meeting all necessary key metrics concurrently—such as high remnant polarization (2*Pr*), high write endurance under continuous bias stressing, large and stable memory window (MW), non-volatile retention, temperature stability, low operating voltages and high breakdown field strength—requires further material and device engineering.

Several approaches have been proposed to mitigate challenges in HZO, including doping[20], seed and capping layer integration[21], changing stack consistency[22] and annealing conditions[23]. However, the reduction of interfacial defects states is difficult to control. Conventional fabrication processes often require vacuum breaks between nitride and oxide depositions, an essential step for HZO capacitors with TiN electrodes. Naturally, the exposure to air leads to carbon incorporation at interfaces, uncontrolled oxidation of the TiN layer, and other interfacial impurities that can eventually act as charge trapping centers. These defects degrade ferroelectric properties, limiting the stability of the MW, endurance, and intermediate state reliability due to unpredictable domain pinning centers[24]. In this work, we demonstrate results of a fully vacuum-integrated stack growth process using an ALD cluster tool that eliminates ambience exposure between layers, preserving pristine interfaces. A HZO layer with 10 nm thickness is grown on top of 30 nm TiN and treated with different annealing conditions to optimize the performance as a multi-bit memory, the schematic of which is shown in Fig. 1a. Electrical characterization of the deposited ferroelectric capacitors (FeCAPs) reveals enhanced ferroelectric polarization, and a high oxide breakdown (BD) voltage for devices annealed at 600 °C. The combination of these two properties enables the use of 32 well-separated stable analog states, albeit at the expense of an increased thermal budget. The material stack is characterized structurally with grazing

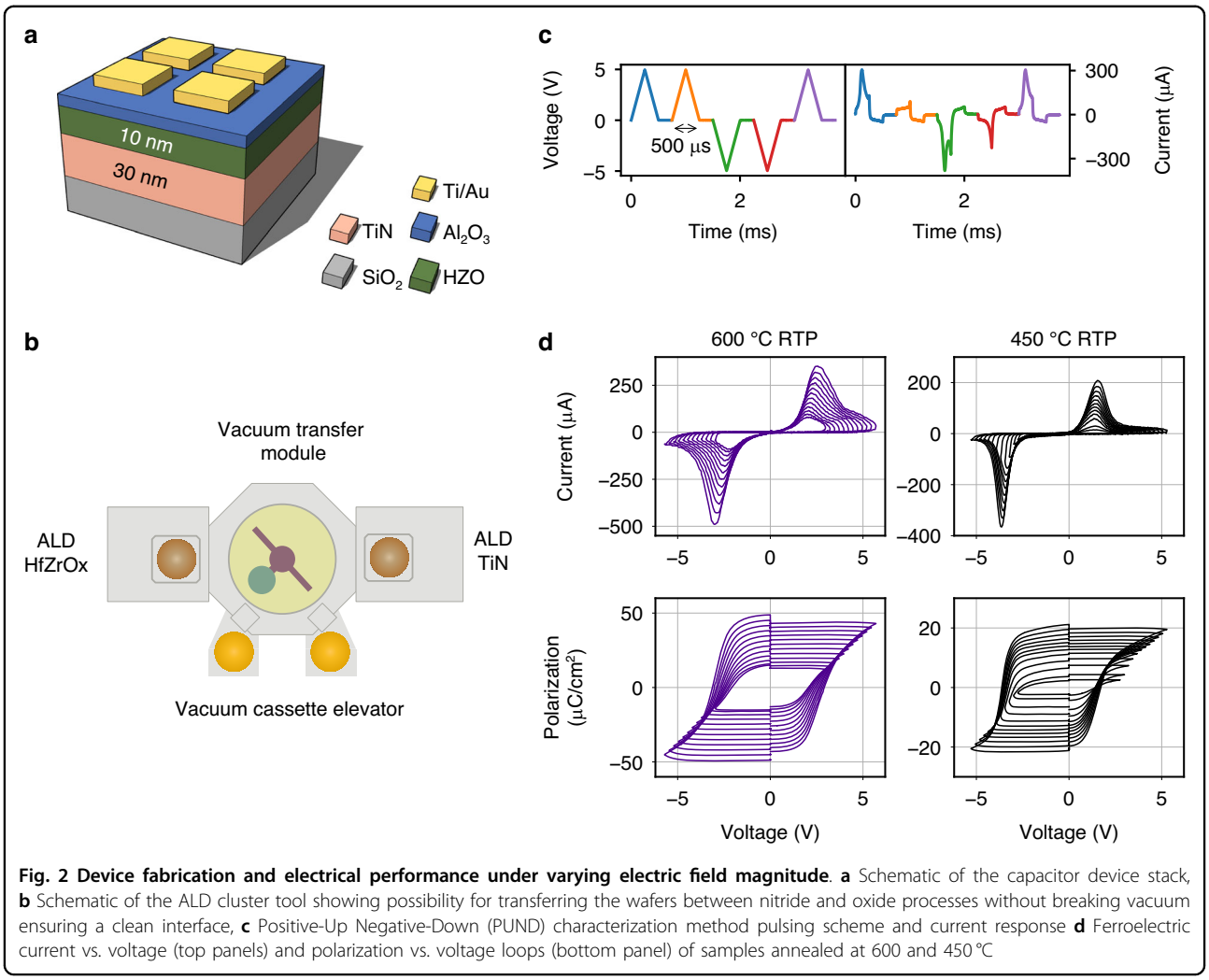

Fig. 2 Device fabrication and electrical performance under varying electric field magnitude. **a** Schematic of the capacitor device stack, **b** Schematic of the ALD cluster tool showing possibility for transferring the wafers between nitride and oxide processes without breaking vacuum ensuring a clean interface, **c** Positive-Up Negative-Down (PUND) characterization method pulsing scheme and current response **d** Ferroelectric current vs. voltage (top panels) and polarization vs. voltage loops (bottom panel) of samples annealed at 600 and 450 °C

incidence x-ray diffraction (GI-XRD), transmission electron microscopy (TEM) and x-ray photoelectron spectroscopy (XPS), which reveals minimal differences in material properties caused by the two annealing conditions. The observed 32 states are then leveraged to perform capacitive IMC tasks, and MAC operation results are compared with different bit resolutions. Expanding on the MAC operation accuracy, the performance of the capacitors is assessed in a hyperdimensional computing (HDC) task. In HDC, input data is encoded into a high-dimensional space, called hypervectors (HVs), which are then used for learning through algebraic operations[19,25–27]. Data manipulation using HVs is an example use-case for IMC, as mathematically simple operations like large vector-matrix multiplications (VMM) are computationally heavy with conventional von Neumann computing architectures. HDC is also highly robust against errors[26], making it pair well together with emerging memories, where variability can be relatively high.

To evaluate the benefits of using the characteristics from the fabricated FeCAPs towards HDC, three of these 5-bit capacitors were paired together achieving 15-bit multi-level cells, the schematic which is shown in Fig. 1b. An inference accuracy of 92.3% is achieved, which is higher and occupies 2.3x less area than binary implementations. Our results show that FeCAP based HDC is accuracy-wise comparable to graphics processing unit (GPU) and ferroelectric field-effect transistor (FeFET)-based computing[25], but with additional benefits from the capacitive computing, including smaller total power consumption[11], denser integration capabilities[12], and a higher number of available bits.

Results

A. Electrical properties

The Metal-Ferroelectric-Insulator-Metal (MFIM) FeCAP stack used in this work, consisting of TiN/HZO/Al_2O_3/Ti/Au is shown in Fig. 2a. The TiN (30 nm) bottom

electrode, HZO (10 nm) and Al_2O_3 (1.2 nm) layers were deposited with the Applied™ Picosun™ single wafer cluster ALD system, illustrated in Fig. 2b. Consequently, no break in the vacuum condition occurred between the bottom electrode nitride and oxide depositions. The ferroelectric crystal phase forms during a rapid thermal processing (RTP) step, and the impact of the RTP temperature is studied with two samples, annealed at 600 C and 450 °C for 30 s under N_2 atmosphere. Choice of the annealing temperatures is motivated by CMOS BEOL compatibility factor[13] (450 °C) and to test what is achievable with higher thermal budget (600 °C). The devices ferroelectric properties were characterized by a positive-up negative-down (PUND) method, with the applied pulse shape and corresponding raw current response shown in Fig. 2c. Subtracting the non-switching current response (corresponding to the second pulse of same polarity) from the total switching current response (corresponding to the first pulse of same polarity), the purely ferroelectric response is extracted. Details on the fabrication, characterization methods and tools are in the section "Materials and Methods". Top and bottom panels of Fig. 2d respectively show the ferroelectric current vs. voltage (I–V) plot and the polarization vs. voltage (P–V) loops for each sample, obtained with the PUND method. Both samples, annealed at 600 °C and 450 °C, show robust ferroelectric switching properties. With increasing voltage amplitude, the P–V loops become wider, leading to higher remanent polarization ($2P_r$) and larger MW. The peaks in the dynamic I–V curves correspond to the switching of FE domains near the coercive voltage (V_C) of HZO. Higher voltages lead to larger peak currents, reflecting increased ferroelectric domain switching.

The annealing temperatures chosen are widely adopted for HZO, and it has been reported that a higher processing temperature tends to increase the ratio of the ferroelectric orthorhombic phase over non-ferroelectric phases[13]. Consistently, improved ferroelectric behavior is observed here when annealing at 600 °C instead of 450 °C with increased $2P_r$ from 22 μC/cm^2 into 32 μC/cm^2 at a 4 V write voltage. The largest observed difference between the samples arises from the high breakdown voltage of the 600 °C sample, which could sustain voltages up to 5.7 V reliably, while the 450 °C sample suffered oxide breakdown at 5.2 V. At voltages of over 5.5 V, the 600 °C sample reached high $2P_r$ values of over 70 μC/cm^2.

Highest polarization values reported so far with a similar thickness of HZO and voltage scale are in the range of 60–65 μC/cm^2[20,28,29], and are achieved with advanced electrode engineering. In this work, we show that increasing the breakdown field of the material to achieve high polarization is possible with thermal engineering and by avoiding additional vacuum breaks between the bottom electrode and functional material deposition. Ambient exposure introduces carbon contamination and spontaneous oxidation of the TiN layer, which affect the electrical and structural properties of the material[30]. These defect inclusions can degrade ferroelectric properties, including intermediate state reliability due to unpredictable domain pinning centers and depolarization fields[24].

Importantly, the BEOL processing compatibility, for monolithic 3D integration of memory with CMOS readout, requires maintaining fabrication temperatures below 500 °C[13], imposing serious limitations on the higher temperature annealing. These devices could, nevertheless, be used in a heterogeneous integration[31]. While the remanent polarization of the 450 °C sample is not as high, it can still provide a high number of analog states, albeit less than the 600 °C annealed sample. The number of states alone does not quantify the analog programmability towards IMC applications, but rather the distribution and overlap of them defines the performance, discussed in following sections.

Analog programmability is of high importance for IMC applications, and the high polarization and breakdown voltages of the samples yield multiple well-defined intermediate polarization states. Using programming pulses of equal width while increasing the voltage amplitude, we extract 32 well-separated states from the 600 °C annealed device. The states are illustrated in Fig. 3a, with Gaussian fits corresponding to each write voltage, obtained from 10 consecutive pulses. The Gaussian model is based on means and standard distributions of the polarization state obtained with each write voltage, which is visualized in Supplementary Information Fig. S1, where the numerical values of the parameters are shown for both the samples. Equal triangular pulses with a risetime of 250 μs were used in all cases, and each state is separated by 0.1 V. Similar extraction was done for the 450 °C annealed device in Fig. 3b, but in this case the number of available states is limited to 16, due to the reduced breakdown (BD) voltage. Noticeably, for both samples the polarization states tend to cluster at lower voltages, while a more linear, well separated response is observed at higher voltages. This behavior is illustrated in Fig. 3c and d, where the symmetry and linearity are the focus from the potentiation and depression characteristics. Some additional stochasticity arising from non-uniform crystallization during annealing can be observed from variations between devices, which is shown in Supplementary Information Fig. S2.

Since polarization itself is difficult to utilize for in-memory computing purposes, a more accessible polarization-modulated quantity is typically used for the computing. In ferroelectric capacitors, the polarization state of the device impacts the capacitance of the system. This can be observed in Fig. 4a, where the characteristic

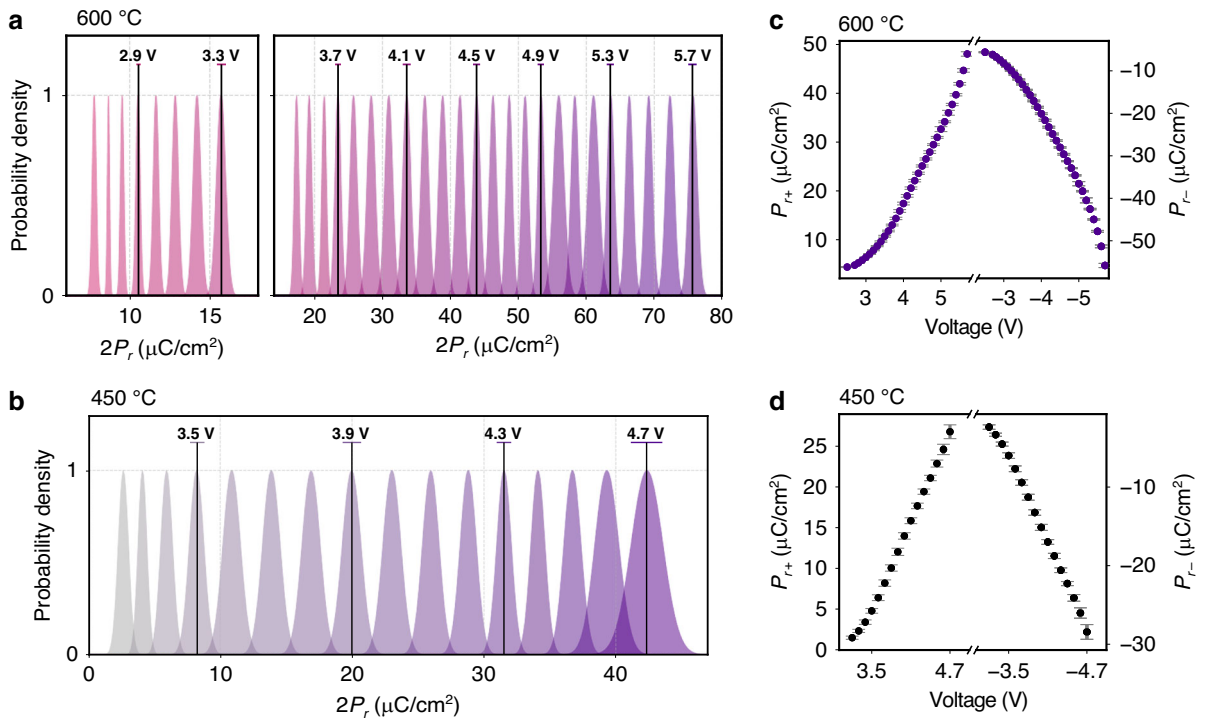

Fig. 3 Analog state operation reliability and linearity. Distributions of the 32 and 16 analog states in the 600 °C and 450 °C device, respectively, with a 0.1 V write voltage separation. **c**, **d** Symmetry and linearity of analog states in both potentiation and depression for the 600 °C and 450 °C annealed devices. The error bars indicate the standard deviation in 10 consecutive pulses

butterfly-shaped curve for *C-V* is obtained in different polarization states for the device annealed at 600 °C. The curves do not exhibit symmetric behavior with respect to positive and negative bias polarities which is a common feature of FeCAPs with asymmetric electrodes. The capacitance value observed at zero voltage is extracted and plotted against the difference in polarization in Fig. 4b. Notably, the response is highly linear, and a first order polynomial was fitted to produce an analytical law between the polarization and the observed capacitance at zero applied voltage. For $V = 0$, a relationship of $C(\mathrm{pF}) = -0.48\left(\frac{\mu m^2}{cV}\right) \cdot P(\frac{\mu C}{cm^2}) + C_0(\mathrm{pF})$ is obtained, with the constant value of $C_0 = 872$ pF. The linear change in capacitance with polarization indicates a high write programmability for capacitive in-memory computing. Moreover, the read voltage can be adjusted to modify the difference in capacitance caused by the polarization states, although the choice of the read voltage also impacts the linearity.

The impact of the selected read voltage on the measured capacitance is shown in Fig. 4c, while the corresponding linearity metrics—namely the R^2-score and the root mean squared error (RMSE), extracted from the linear fit to the data as in Fig. 4c—are plotted in Fig. 4d. The results reveal an optimal voltage window to be between −0.3 V and 0.5 V. Although read voltages between 0.5 V and 1 V yield excellent linearity through the R^2-score, they are disregarded due to the higher RMSE values. Furthermore, the selected read voltage must be sufficiently low not to impact the polarization state of the ferroelectric device. This requirement is favorable from both the device reliability perspective and the standpoint of overall system energy consumption. These results demonstrate that a low read voltage can be used to achieve a non-destructive readout.

B. Structural analysis

Structural analysis for both the 600 °C and 450 °C annealed samples was carried out using different techniques to understand the impact of the thermal engineering on the electrical properties of the FeCAPs. Figure 5a shows the GI-XRD patterns of the HZO/TiN stack in 2θ scan range of 10° to 80° in the insets. In the main figures, the high-resolution short range 2θ scans around 31° is shown. Results show that the material has coexistence of the non-ferroelectric monoclinic (*m*-phase), as well the ferroelectric tetragonal (*t*-phase) and orthorhombic phases (*o*-phase). By deconvoluting the peak around 30.7°, the phase ratio was calculated, and is shown in the legend of Fig. 5a. The raw diffraction peak was deconvoluted to contributions from (*m*-111), (*m*111) and the highly overlapping *o*- and *t*-phases, and with peak centers at

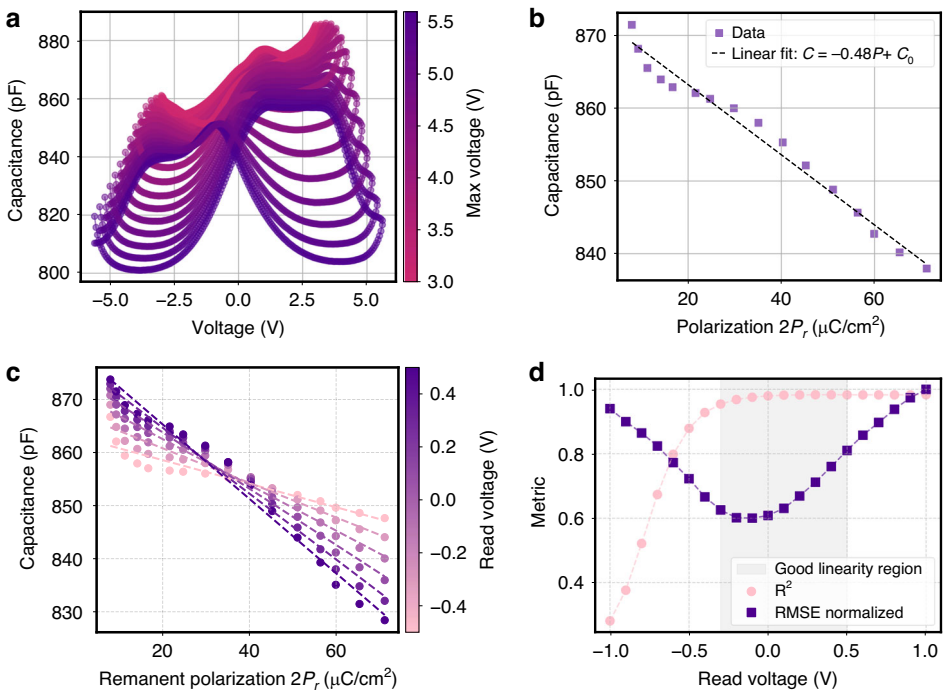

Fig. 4 Capacitive properties modulated by the ferroelectric polarization. **a** The capacitance–voltage (*C* –*V*) characteristics of the 600 °C annealed device measured with 1 kHz frequency with different WRITE voltages. **b** Extracted zero-voltage capacitance (returning sweep) corresponding to different polarization values indicates a linear relationship between the capacitance and polarization values. **c** Linearity of the written polarization states depending on the used read voltage. **d** Linearity metrics analyzed for a larger sweep of read voltages. R^2-value should be as close to 1 as possible, and RMSE as low as possible. The best linearity of the states is identified in the read range between −0.3 and 0.5 V

28.4°, 31.7°, and 30.7°, respectively. The differences in phase ratio between the two investigated samples are not major. However, about a 2% increase in the total ferroelectric *t*- and *o*-phase is seen in the 600 °C annealed sample. Overall, the ratio of the ferroelectric phases over the non-ferroelectric *m*-phase is high in both the samples, which can lead to the high polarization charge density seen in these devices. Both the samples exhibited wake-up free hysteresis, which indicates a high *o*-phase content.

Beside the GI-XRD suggested crystal phase ratio, little difference was structurally found between the two samples. Hence, we will focus our analysis on the 600 °C annealed sample. Figure 5b shows cross-sectional TEM results from the film showing sharply defined interfaces between HZO and TiN together with crystalline phase of HZO layer.

In Fig. 5c, an XPS depth profiling is shown from O 1s and C 1s. Environmental C peaks are known to dominate any thin film surface once the vacuum condition of a sample is broken. To check the cleanliness of the interface in the samples, depth profiling through C 1s of the entire stack was conducted. No C 1s peak at the top and bottom HZO interface was found confirming cleanliness of the interface fabricated in the cluster tool. From the O 1s spectra, the center of the oxygen peak around 530 eV is asymmetric and shifting, which suggests different oxygen bonds are prevalent within HZO.

To resolve different oxygen species, the O 1s spectrum was deconvoluted at each depth. The three components are represented as 1) the stoichiometric Hf-O/Zr-O component at a lower binding energy (~529–530 eV), 2) oxygen vacancies or loosely bonded surface oxygen at intermediate energies (~530–531 eV), and 3) surface hydroxyls (-OH) or chemisorbed water at higher binding energies (~531–532 eV)[32]. The resulting deconvolution peak intensities of the depth profiling (Fig. 5d) show that moving from the top to the bottom surface of HZO, the O 1s shifts towards a lower binding energy. This confirms the presence of -OH groups near the top surface, while in the bulk of the 10 nm film, a more stoichiometric composition is observed, as seen from the individual decompositions in Fig. 5e. The exact peak positions and contributions depend on the Hf content of the film, processing conditions, and the presence of other species like carbon or hydrogen.

Oxygen concentration in HZO and at the HZO/TiN interface was also studied by quantitative deconvolution of Hf 4f and Zr 3d spectra into Hf_{4+}, Hf_{2+}, Hf_0, Zr_{4+}, Zr_{2+}, Zr_0 oxidation states. In Fig. 5f, Hf_{4+} and Zr_{4+} are high in the top surface of the film, Hf_{2+} and Zr_{2+} peak in

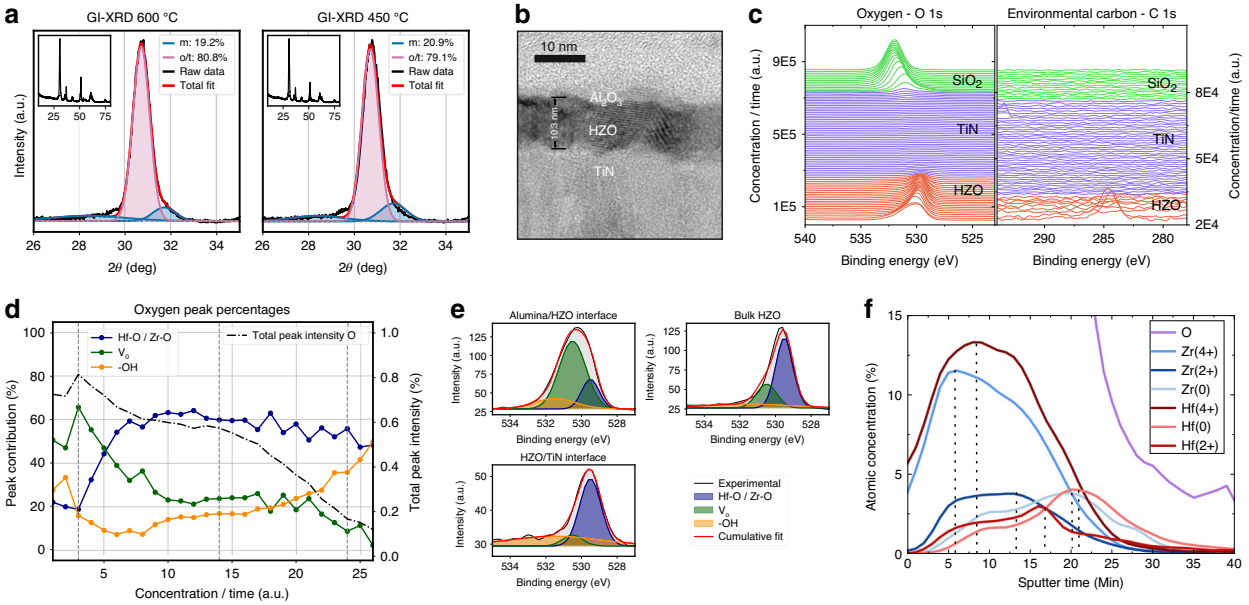

Fig. 5 Structural data confirming the quality of the devices. **a** GI-XRD spectrum peak analysis of $HfZrO_x$ near 31° for estimating orthorhombic (o), tetragonal (t) and monoclinic (m) phase ratio. Inset: broad 2Θ range of 10–80°. **b** TEM image of the ferroelectric capacitor stack showing well-crystallized HZO of the 600 °C annealed sample. **c** XPS depth profile for O 1s and C 1s peaks showing the HZO top and bottom interface, defined by diminishing O 1s peak at TiN and lack of C 1s peak at both these interfaces. **d** Result of O 1s peak deconvolution into metal-oxide, oxygen vacancy (V_O) and hydroxyl group (-OH) with respect to depth. The vertical lines represent the deconvolutions plotted in **e**. The cumulative fits of the oxygen peak deconvolution results at different depths **e f** XPS depth profile of the HZO sample showing oxidation states of Hf and Zr as a function of depth from the top surface

the middle while Hf_0 and Zr_0 peak near the HZO/TiN interface. This indicates a continuous decrease of oxidation states of the Hf and Zr species from surface to bottom interface, likely due to TiN scavenging O_2 from HZO during the RTP step, which is expected in these samples[13].

Overall, the results of Fig. 5d-f point out that the oxygen intensity is highest in the alumina layer, interfacing ambient air. In the alumina/HZO interface, we see a high intensity of oxygen vacancies. In the bulk HZO, the oxidation state of hafnium and zirconium is 4+, which is consistent with the higher metal-O peak. Nearing the HZO/TiN surface, the oxidation states of the Hf and Zr start decreasing, indicating metallic hafnium and zirconium. Simultaneously, the total oxygen concentration is decreasing together with the oxygen vacancy intensity, while the surface hydroxyl group or loosely bonded surface oxygen is gaining prevalence. Likely, near the TiN, oxygen is scavenged from the Hf and Zr, forming a titanium oxide dead layer. Instead of leaving behind oxygen vacancies, -OH groups or surface oxygen is formed. There are trace amounts of oxygen also inside the TiN, however, the total intensity is below 5% of the maximum oxygen intensity.

This analysis suggests that in the pristine state, oxygen vacancies are concentrated around the Alumina/HZO interface, instead of the HZO/TiN interface. The impact of the annealing conditions was minimal in the XPS, and the two compared samples showed near-identical oxygen deconvolution results (Supplementary Information Fig. S3). The two samples exhibited different combined *o/t* to *m* phase ratios, but the 2% difference in the ratio alone does not explain the large distinctive polarization and breakdown characteristics. Since oxygen vacancies are considered the most important electron trap in polycrystalline HZO—typically responsible for the hard breakdown at high voltages (>5.7 V)[33]—their properties and distribution are further analyzed in section "Device Reliability" using device simulations performed with the Ginestra™ software[34].

C. Capacitive in-memory computing

With the rise of neural networks in machine learning, capacitive in-memory computing is becoming increasingly attractive for essential operations such as high-throughput matrix multiplication. Analog MAC operations are often implemented on hardware in the form of crossbar arrays. A comparison of different emerging memory technologies is presented in Table 1: resistive random-access memory (RRAM), phase-change memory (PCM), magnetic random-access memory (MRAM), ferroelectric random-access memory (FRAM) and FeCAP based memory. While being fast and energy efficient,

Table 1 Quantitative comparison of different emerging IMC technologies. Reproduced with permission from ref. [2]

Parameter	RRAM-based IMC	PCM-based IMC	MRAM-based IMC	FRAM-based IMC	FeCAP-based IMC (Ferroelectric capacitor crossbar)
Static power	~50–200 pJ per 128 × 128 MAC (inference)[9]	Negligible (few eV)[49]	No static power; leakage only through transistors[50]	Negligible static power[11]	~3.8 pJ per MAC (128 × 128 array)[9]
Sneak-path currents	Significant; can exceed selected cell's current[51,52]	Present without selectors[49]	Eliminated by 1T-1MTJ structure[53]	No sneak paths[11]	None; inherently selectorless[10]
IR-drop (voltage drop)	~7.7 mV across 256 cells at 1.7 V (~4.5% loss)[54]	Notable in large arrays[55]	Present during fully parallel read[53]	High inherent cell resistance; negligible drop[11]	Minimal due to negligible steady-state current and high impedance[56]
Read disturb	Requires ~ 0.1–0.2 V for nondestructive read[57]	Negligible; nondestructive[55]	$>10^9$ disturb-free reads[53]	Nondestructive read at low voltages[11]	Nondestructive, near-zero small-signal read voltage[10]
Endurance (write cycles)	10^5–10^7 typically; up to 10^9–10^{12} in lab tests[57]	10^6–10^8 cycles[58]	Extremely high; $>10^{12}$–10^{15} cycles[58]	10^{10}–10^{15} cycles[58]	10^{10}–10^{12} cycles for HZO-based FeCAP[11]
Selector requirements	1 T1R or 1S1R configurations; self-rectifying cells partially mitigate[52]	Requires selector or 1T1R[59]	Inherent 1T-1MTJ structure[53]	Integrated selector via FeFET structure[11]	No separate selector required[11]
3D stacking	Possible via self-rectifying arrays, at energy cost[60]	Feasible; demonstrated[61]	No commercial 3D MRAM yet; research ongoing[62]	Dual-layer 3D FeRAM (32 Gb) demonstrated[58]	BEOL-compatible; two-terminal structure enables vertical stacking[10]
Small-signal read voltage	~0.1 V[63]	~0.1–0.3 V[64]	~50–200 mV[50]	~0.8–1.0 V (FeFET threshold sensing)[11]	Few millivolts; much lower than coercive voltage[10]
Analog linearity	4–6 bits per cell practical[65]	Nonlinear conductance; mitigated by cell averaging[64]	Limited by tunneling magnetoresistance (TMR); binary nature dominates[50]	Hysteresis and saturation affect analog linearity[11]	Linear charge accumulation; limited dynamic range (~2 bits)[11]
Energy per MAC	100–200 fJ/MAC (130 nm CMOS)[66]	~0.1 pJ/MAC (10–12 TOPS/W)[64]	~0.04–0.1 pJ/MAC[50]	885.4 TOPS/W for 2-bit MAC ($\sim 1 \times 10^{-15}$ J per operation)[67]	Sub-10 fJ/MAC achievable; sub-femtojoule regime projected[11]

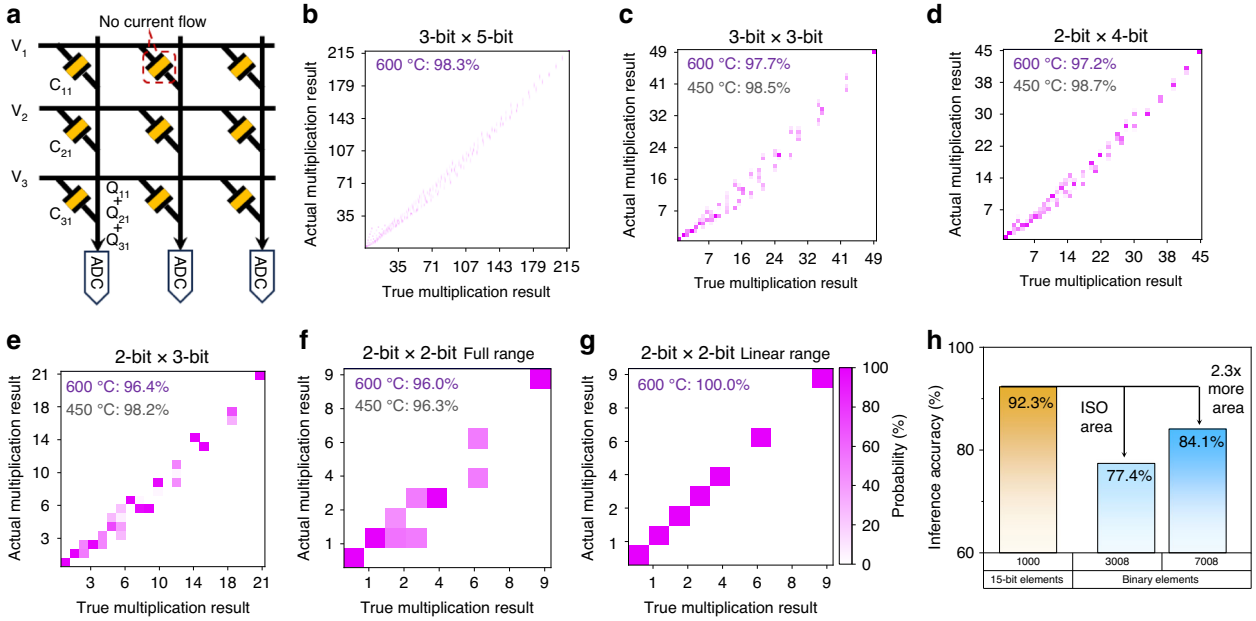

Fig. 6 Structure and output from the simulated in-memory computing. **a** Structure of a charge-domain crossbar using FeCAPs. Each FeCAP stores a value with is polarization modulating the capacitance which is multiplied by the applied input voltage. The resulting charges are accumulated on the shared wire and digitized with and analog-to-digital converter (ADC). Variability analysis of analog multiplication with 600 °C annealed devices, with a linearity scores associated for both annealing conditions. The confusion matrices show the probability distribution for different analog input resolutions (voltage resolution × FeCAP resolution) with **b** 3-bit × 5-bit, **c** 3-bit × 3-bit, **d** 2-bit × 4-bit, and **e** 2-bit × 3-bit. **f** Confusion matrix with 2-bit × 2-bit input resolution using the full write voltage range of the FeCAP. **g** Utilizing the strong linear state range leads to exact and correct multiplication result. **h** Inference accuracy of HDC using random projection encoding with 15-bit values vs. binary encoding for the ISOLET dataset. Three 5-bit FeCAPs are used to store one 15-bit element, i.e., 3x more storage devices for the same dimension compared to binary storage devices

binary crossbars consume a significant amount of costly chip area to store the weights, a challenge that becomes increasingly severe as models continue to grow. Hence, multi-bit storage within a single device is highly sought after, as it increases storage capacity with maintaining similar area requirements. Enhanced bit density is therefore crucial for enabling larger models to be stored on-chip, or for minimizing data transfer to off-chip memory.

Besides enabling multi-bit storage, a charge-domain computing scheme has several advantages over classical current-domain approaches. In current-domain schemes, the variation and non-linear output characteristics pose significant challenges, which can be mitigated in charge-domain[3]. Additionally, as no continuous current flows through the capacitor during computation (see Fig. 6a), active power consumption is greatly reduced, improving overall energy efficiency. Specifically, FeCAP-based IMC has been shown to be highly energy efficient compared to other non-volatile memories[7].

Computing accuracy

In the multi-level FeCAP presented in this work, synaptic weights are encoded in the polarization states, which modulate the capacitance, C, as seen in Fig. 4. During computation, the analog input voltage (V) is multiplied by the capacitance, producing a proportional charge $Q = V \times C$. To show the IMC capabilities of the fabricated FeCAPs, analysis was performed using experimentally calibrated polarization state distributions, shown in Fig. 3a and b. Voltage capacitance products were sampled using Monte Carlo methods and digitized with an analog-to-digital converter using equidistant decision boundaries to populate the probability distributions.

The VMM capability of a capacitive crossbar array with all 32 states used as 5-bit resolution is shown in Fig. 6b as a confusion matrix where the mathematically correct multiplication results are on the diagonal. The matrix multiplication accuracy score is defined in section "Materials and methods". The multiplication is done using the analog voltage levels to encode the first operand, and multi-level FeCAP states to store the second operand. The results point that the inherent variation of the 32 states and asymmetric charge contribute to errors in the actual output of the crossbar device. These errors can be mitigated with reducing input bit resolution, which can be seen in Fig. 6b–e, where the accuracy scores obtained from the 450 °C annealed devices with 16 available states are also shown. The corresponding confusion matrices are

found in Supplementary Information Fig. S4. The computational accuracy of the matrix multiplication is a trade-off, where more accurate multiplication results are achieved with decreasing memory density. With reduced input resolutions, the probability distributions tighten due to wider margins, reducing errors.

Reducing the bit resolution alone might not increase the multiplication accuracy, since the asymmetric nature of charge updates also contributes to incorrect multiplication results. Notably, the accuracy scores computed from the 450 °C devices are consistently higher than the corresponding result from 600 °C devices. This is promising for the lower thermal budget requirements of 3D monolithic integration of HZO FeCAPs with CMOS readout, since the linear range is reached within a lower write voltage. Furthermore, Fig. 6f shows a 2-bit × 2-bit input resolution used with the full write range of the FeCAP, still producing variation from the mathematically correct result. Figure 3c reveals a stronger state linearity achieved at voltages above 3.8 V, where a 2-bit resolution is still achievable. Figure 6g shows that limiting the states to this stronger linearity yields mathematically correct multiplication results without observing deviations, which was only possible with the 600 °C devices. This can be attributed to the more consistent standard deviations across different voltages and better linearity from the 600 °C devices, as is shown in Supplementary Information Fig. S1. The result highlights the importance of linearity and tight-state distributions to achieve exact results, but within systems with higher error resilience, the full 32 available states can be utilized as 5-bit memories.

Brain-inspired HDC is an alternative to classic deep learning, posing strong error resilience[26], essential for the usage of emerging memories. We evaluated the classification performance on the ISOLET dataset[35], consisting of spoken letters, with the random projection encoding scheme[36] and iterative training over 10 epochs. Three multi-level cell FeCAPs store one 15-bit number. With 15-bit fixed-point numbers and a dimension of 1000 elements per vector, we achieved an inference accuracy of 92.3%, as illustrated in Fig. 6h. To highlight the effectiveness of high bit-density, we compared it to classical binary encoding (BE). For a BE system occupying the same area, the vector dimension is 3008 elements per vector, and the corresponding inference accuracy is much lower at 77.4%. Increasing the vector dimension to 7008, the BE results in an accuracy of 84.1%, while occupying 2.3× more area. This shows that for error resilient tasks such as HDC, these 5-bit FeCAPs provide superior capabilities over conventional binary encoding.

Hyperdimensional computing with ferroelectric in-memory computing has been explored before with FeFETs[24], with the accuracy on the ISOLET dataset reaching as high as 96% with a 3-bit FeFET, corresponding to the same accuracy as obtained with a graphical processing unit (GPU) implementation. However, differences in encoding methods, training parameters and vector dimensions can impact the inference accuracy[26]. The number of dimensions per vector can be decreased with increased bit-per-cell number while maintaining accuracy[25], as supported by the results in this work. It has been shown that FeFETs can achieve higher accuracies than FeCAPs for IMC tasks but with a higher total energy consumption and compromised reliability[11]. We showed that with the inference accuracy of 92.3%, the ferroelectric capacitors presented in this work are competitive against FeFETs in HDC tasks with decreased power consumption. In addition, the FeCAP based approach benefits from the dense scalability of two-terminal devices over the three-terminal FeFETs[12].

Energy efficiency and non-idealities

Important characteristics for IMC tasks are energy consumptions of the write and read operations. Write energy is the hysteretic energy needed in the rewriting of polarization. With multi-bit elements, writing happens from one state to another, and the largest difference for these devices is from −5.7 to 5.7 V, which will lead into a write energy of $E_{\text{write}} = A \cdot \int_{-P}^{P} V(P)dP \approx 200\text{nJ}$ per one write event. For comparison, a binary implementation with the same capacitors, a lower write voltage can be used. To obtain a $2P_r$ of 50 μC/cm^2, the required write energy would be in the range of 100 nJ, which would need to be written to 5 devices to carry the same information, leading into a total write energy of 500 nJ for the binary case. So, using 5-bit elements is energetically more efficient even when writing from the minimum to the maximum.

The energy required to read the memory state from the capacitors is $E_{\text{read}} = \frac{1}{2}CV^2$. The energy is highly sensitive to chosen read voltages. In the present case, it is identified that read voltages should not exceed 0.5 V from the linearity point of view of the capacitors, however, there are no known device level limitations to how low the read voltages can be, as is mentioned in Table 1. For example, with a read voltage of 0.1 V and the capacitance from Fig. 4b, $E_{\text{read}} \approx 4.3\,\text{pJ}$, which is much lower compared to the write energy. The practical limitations of the read voltage arise from the analog to digital conversion. The read energy of an individual device is not highly sensitive to the binary or 5-bit encoding but the same information can require five times more read events in the binary case than with 5-bit individual devices.

Both the read and write energy are linearly area dependent. The area of the large capacitors used in this work are 200 μm × 200 μm = 40000 μm^2. To highlight the importance of device scaling to reach the optimal energy benefits of capacitive computing, by bringing the feature

size down to $2\,\mu m \times 2\,\mu m = 4\,\mu m^2$, the read energy will be brought down significantly to $E_{read,2um} \approx 0.43\,fJ$. With this same scaling, the write energy will be scaled to $E_{write,2um} \approx 20\,pJ$. However, with the scaling of the FeCAP elements, parasitic capacitances start to become non-negligible in the array impacting the accuracy of computation. This is a non-trivial topic and needs to be further explored in future work.

We note that other important nonidealities to consider are latency and parasitic effects. Latency in the circuit will arise from the required N-bit Successive Approximation Register (SAR) and Analog to Digital Converter (ADC), rather from the device side. Parasitic IR drop should not be an issue, since no current is flowing through the capacitor in the compute step[2]. Impact of parasitic capacitances are non-trivial for charge-based in-memory computing, that is dependent on the wiring of the circuit, however it is expected to be much lower than the capacitance of the large capacitors used in this work.

FeCAP as storage device

Using the proposed FeCAP as a storage device, we analyze the impact of the state variation on the inference accuracy of the shown Hyperdimensional Computing application. For this, we use the extracted variation data shown in Fig. 3a. We place 32 states for a 5-bit storage device, equidistantly spaced within the extended memory window of the 600 °C annealed FeCAP, and compute the error probabilities for each state. This is done by intersecting the normal distribution of each state with its direct neighbor. The list of intersection points forms the decision boundaries between the states and quantizes polarization values to their respective state. Given the decision boundaries and the normal distributions, a confusion matrix is computed that describes the probability of each state being misinterpreted as a neighboring state due to overlaps. For the 5-bit FeCAP, Fig. S5a show the error probability for each state with a maximum error of 4.2%.

To study the impact of variation, we inject the errors described by the confusion matrix as bit flips into the class or sample hypervectors of the ISOLET dataset using the fixed-point HDC algorithm. Here, each vector element is a 15-bit fixed-point float number that can be split and stored in three 5-bit FeCAPs. To inject the bit flips into the vectors, each element's bit string is cut into the blocks stored by each FeCAP (e.g., 5-bit blocks). Per block, the bit pattern's associated state is randomized using the discrete probability distribution described by the confusion matrix and then shifted back to the original position in the floating-point number. This means that errors in the states of FeCAPs that store the most significant bit will have the greatest impact on the error in the value of the number.

Using this error injection technique, we conduct three studies and inject the errors in the vectors that store the 1) class vectors (i.e., the model), 2) test vectors, or 3) at both places. To gather statistical data on how the errors will impact the inference accuracy, the full inference run with all test vectors is repeated 100 times per study. Fig. S5b shows the resulting inference accuracy distributions with the ideal inference accuracy of 92.3% indicated as the dashed line. Injecting the error into the test vectors maintains accuracy with deviations below ±0.5 percentage points. When injecting errors into the class vectors, the mean accuracy drops by ~6 percentage points to 86%. Errors in both the class and test vectors further reduce inference accuracy to 83.3%, a 9-percentage-point drop. This shows that sensitivity to errors depends on which data are affected.

To improve the error rate, and for example to store the sensitive class vectors more reliably, the bit density can be reduced to 4 bits per FeCAP *(i.e.*, 16 states), which yields a maximum error probability of 3.7e-4%. This low error probability imposes little deviation on inference accuracy, and thus, the few class vectors can be stored in lower-bit-density FeCAPs, while the test vectors are stored at a higher bit density in 5-bit FeCAPs.

D. Device reliability over continuous bias stressing

The previous analysis of the IMC array pointed out that higher voltage operation is needed for increased number of analog states and linearity, improving the mathematical performance of the MAC operations. Applying higher operating voltages in HZO-based ferroelectric capacitors is challenging from a device perspective, as increased electric fields exacerbate degradation mechanisms such as oxygen vacancy drift and lead to elevated leakage currents, which necessitates thorough testing to determine stable operating conditions. A series of electrical tests were performed on the 600 °C sample to verify accurate device functionality over time despite degradation from multiple write-erase cycles, prolonged bias stressing and internal degradation mechanisms like trap and defect generation and redistribution.

A series of electrical tests were performed on the 600 °C sample to verify accurate device functionality over time despite degradation from multiple write-erase cycles, prolonged bias stressing and internal degradation mechanisms like trap and defect generation and redistribution.

Write-endurance tests investigate the evolution of $2P_r$ over repeated write and read cycles, the results of which are shown in Fig. 7a–b, for the polarization and corresponding coercive voltage. As shown in the inset of Fig. 7b, devices were pulsed with rectangular pulses of ±3 V at 100 kHz frequency. In every 1/3rd of a decade, one dynamic hysteresis measurement (DHM) was performed with triangular pulses of amplitudes 5.7 V, 5 V and

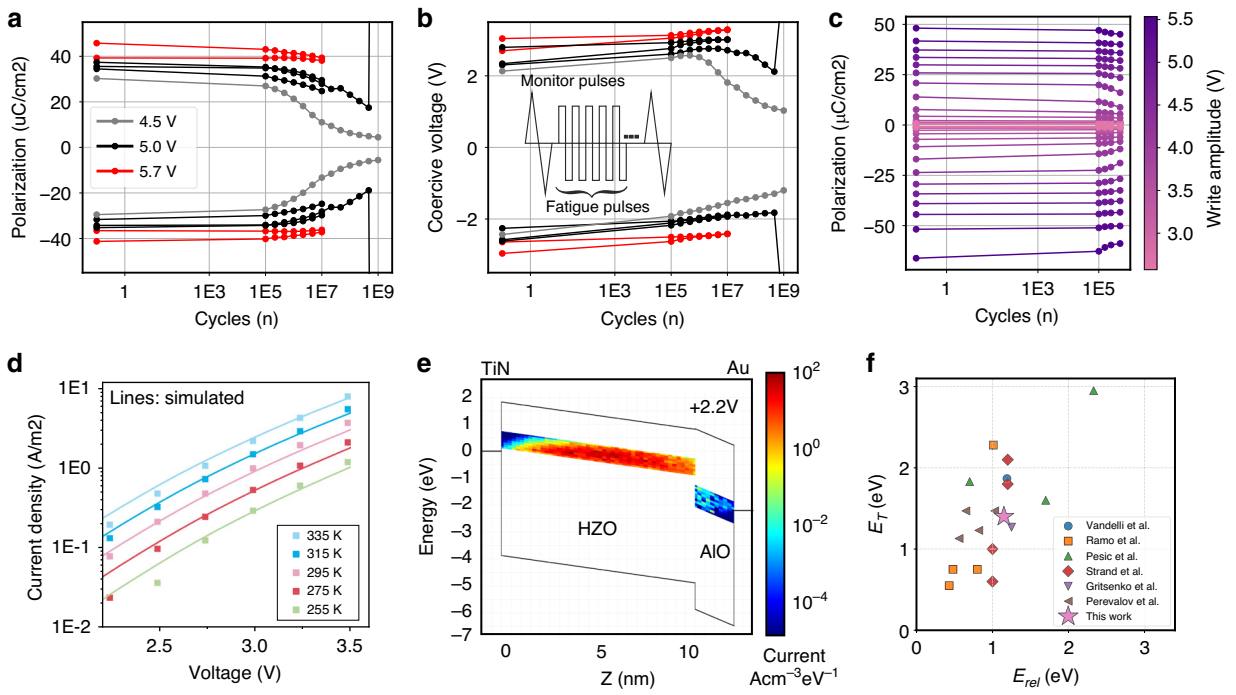

Fig. 7 Device reliability and defect characterization. **a** Endurance of the 600 °C devices under different applied pulse amplitudes showing fatigue effect can be significant when relatively smaller pulse amplitudes are applied. The devices can be recovered from fatigue states. **b** Effect of electric field cycling to the imprint of the devices. **c**. **d** Leakage current density measured (symbols) and simulated (lines) at different temperatures on the TiN/HZO/Al_2O_3/Au device under study. **e** Traps distribution of the 600 °C sample extracted from the simulation of the *I–V* characteristics. **f** Comparison of extracted traps energies with reports in the literature indicate oxygen vacancies as the atomic structure defects in the HZO film

4.5 V with 1 kHz frequency. Devices endured long pulsing cycles exceeding 10^9 cycles even with a low pulsing frequency of 100 kHz, when DHM was measured at 4.5 V. The samples exhibited signs of fatigue reflected by diminishing $2P_r$ (Fig. 7a). Notably, a larger drop in polarization was observed with lower voltage stress, while the oxide breakdown happened later. This effect has been observed earlier in HZO capacitors as well and is attributed to increased domain pinning effect due to redistribution of oxygen vacancies under continuous pulsing[23].

Figure 7b shows the impact of the electric field cycling on the ferroelectric switching field. The change in imprint due to field cycling indicates oxygen vacancy generation and/or redistribution, which is more significant at lower voltages. The endurance characteristics of the 450 °C annealed device are presented in Supplementary Information Fig. S6. The impact of field cycling for the endurance of the 32 analog states is presented in Fig. 7c. While fatigue effects are observed with continuous field cycling, the underlying linearity is maintained, as shown in Supplementary Information S6. Direct current breakdown voltage was tested using time-dependent dielectric breakdown (TDDB) measurements (Supplementary Information S8). Majority of the 600 °C annealed devices endured 4 V continuous bias stress for over 300 s, which is one of the highest reported values for HZO[37], further highlighting the high breakdown endurance of the sample. High breakdown voltages are reported together with better reliability in polarization uniformity, high $2P_r$, and long endurance[10,15,28].

Leakage currents were simulated using Ginestra™[34], further described in section "Materials and Methods". Temperature-dependent leakage currents measured on the 600 °C stack is well reproduced by simulations, Fig. 7d, allowing characterization of the amount and nature of traps actively supporting the charge transport through the material stack. Evidently, the defects in the HZO film are responsible for the conduction, seen from the extracted distribution of traps shown in Fig. 7e, colored according to the current driven at the example voltage of 2.2 V. Extracted trap parameters (density, N_T, thermal ionization energy, E_T, and relaxation energy, E_{REL}) are reported in Table 2 along with the main material parameters, and compared to the values previously reported for Hf- and Zr-based materials in Fig. 7f[33,38–42]. Extracted E_T and E_{REL} values are consistent with those reported for oxygen vacancies. The extracted density of $\sim 2.5 \times 10^{19}\,cm^{-3}$ is at the lower end of the range reported ($3–4 \times 10^{19}$ to $10^{21}\,cm^{-3}$)[33,38,41,43–45], proving the superior quality of the fabricated HZO film.

Table 2 Extracted trap parameters, with Au work function of 5 eV and TiN 4.42 eV

Parameter	Al_2O_3	HZO
κ	9.3	28
E_G (eV)	6.7	5.7
χ_e (eV)	2.58	2.6
m^*/m_0	0.2	0.25
N_T (cm^{-3})	1.58×10^{18}	2.47×10^{19}
E_T (eV)	2.49 ± 0.4	1.4 ± 0.3
E_{REL} (eV)	1	1.15

Discussion

The performance of the fabricated devices is benchmarked in Table 3 against state-of-the-art results reported in literature[10,15,16,20,21,28,29]. The polarization values of our samples are record-high in the thickness scale near 10 nm, only exceeded by individual reports with nanometer scale lateral device area[15]. In addition, the high polarization values of this work are achieved by a simple and cost-effective fabrication process, without breaking the vacuum between the bottom electrode and ferroelectric layer. Achieving higher ferroelectric crystal lattice phase benefits from a higher temperature annealing process but comes with the additional cost of incompatibility with CMOS back-end-of-line processing. The structural analysis and simulations point towards oxygen vacancy related trap sites close to the HZO/Al_2O_3 interface being of key importance in understanding the high breakdown field of the material.

In Table 3, there are reports on high $2P_r$, but analysis on the system level is lacking. The existing system level considerations of 32 bits devices suffer from low $2P_r$, making it more difficult to distinguish the states from each other. This work considers the full picture from fabrication to structural and defect characterization into system level simulations of their in-memory computing capability. We demonstrate that using methods of capacitive computing, the high $2P_r$ value and linearity of the 600 °C samples increase the accuracy of hyperdimensional computing with decreased area requirements. For the first time, we demonstrated the performance of a capacitive IMC system for HDC applications based on ferroelectric capacitors. We achieved accuracy metrics comparable to 3-bit FeFET and GPU implementations, [25] while offering the benefits of higher energy efficiency of capacitive computing.

Previous works with ferroelectric devices and in-memory computing focused on obtaining binary memories, where the operating voltage is under 1 V. It is important to note that, if a high number of states is required maintaining distinguishability and reproducibility, a higher write voltage is necessary in 10 nm HZO FeCAPs. It was previously shown in our work[46] that constrained ferroelectric domain mobility can lead to finer differences in polarization states when subjected to continuous voltage pulsing, leading to high-resolution and distinctly identifiable intermediate polarization states. Therefore, through proper engineering of ferroelectric domain pinning sites, it is possible to enhance linearity and reduced variance of the analog states leading to improved multistate reliability. In this work, the number of states used in a ferroelectric capacitor is either 32 or 16, capable of 5-bit or 4-bit operation, significantly improving the performance of a hyperdimensional computing system through decreased areal requirements and increased accuracy. Increasing the voltage introduces the downside of decreased performance under continuous bias stressing, but with proper stack and thermal engineering, a balance between the high polarization and endurance characteristics is achievable.

Finally, thermally activated leakage currents and oxygen vacancy movement were identified as key reliability issues for maintaining a high number of analog states reliably. Even without changing the process engineering for the current samples, one possible way to mitigate these issues is reducing the total energy associated with these mechanisms by cooling the devices down for their operation, which can further increase the linearity of the polarization states[47]. As can be seen from Supporting Information Fig. S9, the polarization characteristics of the devices in this work stay mostly unchanged when cooled down to cryogenic environments. It has been shown that FeFET-based HDC performance suffers in cryogenic environments due to increased threshold voltage[19]. From the coercive fields of the FeCAPs in Fig. S9, we can conclude that the read voltage will not increase significantly at 4 K for capacitive IMC, so performance degradation of computing is less probable. This hold promises for capacitive IMC as one promising pathway for cryogenic computing.

Materials and methods

The full MFIM ferroelectric capacitor stack is TiN (30 nm)/HZO (10 nm)/Al_2O_3 (1.2 nm)/Ti (5 nm)/Au (50 nm), deposited on a thermally oxidized silicon wafer. The stack nitride and oxides are thermally deposited using Applied™ Picosun™ single wafer cluster ALD system, with which vacuum was not broken during the deposition until the capping layer of alumina. This is followed by a RTP step and top electrode deposition by e-beam evaporation. The two samples reported here were RTP annealed at 600 °C and 450 °C for 30 s under N_2 atmosphere.

Devices are characterized by PUND method with a Ferroelectric Material tester AixACCT 2000E. The PUND

Table 3 Benchmarking against state-of-the-art results of ferroelectric capacitors

Device architecture	Annealing	Wake-up cycles	$2P_r$ (μC/cm^2)	MW (MV/cm)	Endurance	Presented analog states	Breakdown field (MV/cm)	System level consideration	Reference
β-W/HZO (7 nm)/β-W	450 °C for 1 min	1E3	64	3	5E12 fatigue-free	2	6.3	No	Chen et al.[29]
TiN (polished)/HZO (5:5 superlattice, 10 nm)/TiN	450 °C in forming gas	1E3	60	3	4E12 fatigue-free	2	5.1	No	Zhao et al.[28]
TiN/HZO (10 nm)/TiN	600 °C for 30 s	N/A	14	2.5	N/A	32	N/A	Yes – FeFET in neural network	Oh et al.[16]
Mo/HZO (4 nm)/Mo, nanometer scaled	400 °C for 10 min	No	108	3	1E10	2	7.5	No	Huang et al.[15]
TiN/TiO_2/La: HZO (9 nm)/Nb_2O_5/TiN	N/A	N/A	66.5	3	1E11	2	4.5	No	Popovici et al.[20]
TiN/HZO (9 nm)/TiN/Mo/TiN	450 °C for 30 s	1E3	34	2.2	1E10	8	N/A	No	Wang et al.[21]
TiN/HZO (10 nm)/TiN	450 °C for 30 s	N/A	N/A	N/A	> 1E4	2	N/A	Yes – capacitive computing neural network inference	Hur et al.[10]
TiN/HZO (10 nm)/Al2O3 (1.2 nm)/TiN in cluster tool	600 °C for 30 s	No	1st state: 6 32nd state: 75	1st state: 2 32nd state: 6. Linearly increasing	1E8 @ 5 V monitoring 100 kHz pulsing with 3 V, > 1E9 with 4.5 V monitoring	32	5.7	Yes – Capacitive computing in hyper dimensional computing	This Work 600 °C
TiN/HZO (10 nm)/Al2O3 (1.2 nm)/TiN in cluster tool	450 °C for 30 s	No	45	2.2	> 1E9 with 4 V monitoring	16	5.2	No	This Work 450 °C

pulse shapes, parameters and corresponding device current responses are shown in Fig. 2c. A 1 kHz frequency is used with a top electrode area of 200 µm × 200 µm for each capacitor. With the PUND method, the ferroelectric switching polarization is obtained by applying a positive voltage during the P and U peaks, and a negative voltage during the N and D peaks. As the ferroelectric switching occurs fully during the P and N peaks, the U and D correspond purely to dielectric displacement and leakage response, which can then be subtracted from the P and N peaks, yielding only ferroelectric switching currents. This current is integrated to obtain the area-normalized polarization curves. Other electrical measurements were carried out with the Keysight B1500A Semiconductor Parameter Analyzer. No wake-up pulsing was applied before the measurements were conducted, as the ferroelectric switching curve of a pristine device was fully open without it.

The matrix multiplication accuracy is evaluated with a linearity metric defined with equation

$$\text{Linearity score} = \left(1 - \frac{\sum_{o=0}^{N}\sum_{i=0}^{N}\frac{\text{abs}(o-i)}{N}\cdot P(o,i)}{N+1}\right)\cdot 100,\%$$

where $P(o,i)$ is the probability of obtaining the outcome i when o is the expected and correct outcome, and N is the highest multiplication result of the configuration: for a-bit voltage times b-bit FeCAP, $N = (2^a - 1)\cdot(2^b - 1)$. The value of 100% corresponds to mathematically correct output, and a lower score indicates a less precise mathematical performance. This linearity score is suited for comparing results from the same bit configurations.

Leakage currents were simulated using Ginestra™[34], an advanced semiconductor device simulation tool, developed to model the electrical behavior of both conventional and next-generation electronic devices. Defects are treated as discrete localized states, characterized by two main parameters: the thermal ionization energy (E_T) and the relaxation energy (E_{REL}), which reflect their underlying atomic configuration[38]. The simulator incorporates multiple charge transport and trapping mechanisms, including direct tunneling, Fowler–Nordheim tunneling, and trap-assisted tunneling, based on multi-phonon theory. It also accounts for drift-diffusion through conduction, valence, and defect bands, as well as thermionic emission. The transport equations are solved self-consistently along with Poisson's equation, considering the dynamic charge state and occupancy of traps. Further details on the underlying charge transport model can be found in previous publications[38,43,48].

Acknowledgements
The work gratefully acknowledges financial support from KDT-ARTIC project (No. #101139908). ARCTIC project is supported by the Chips Joint Undertaking funded by European Commission, its members (including top-up funding by Belgium, Austria, Germany, Estonia, Finland, France, Ireland, The Netherlands and Sweden), the Canadian and the Swiss federal governments and Business Finland (Project VETURI - Chip Zero, Grant No. 9971_31_2022).

Author details
[1]Faculty of Information Technology and Communication Sciences, Tampere University, Tampere, Finland. [2]School of Computation, Information and Technology, Techical University of Munich, Munich, Germany. [3]Picosun Oy– an Applied Materials company, Kirkkonummi, Finland. [4]Department of Engineering Sciences and Methods (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy. [5]Applied Materials Inc., Santa Clara, CA, USA

Author contributions
Ella Paasio and Sayani Majumdar wrote and edited the manuscript. Simon Thomann planned and carried out the in-memory computing simulations and wrote the corresponding draft. Ella Paasio, Safdar Muhammad, Andrea Padovani, Hussam Amrouch, Gaurav Thareja, Sayani Majumdar conceived and designed the research. Anika Anu, Ella Paasio, Xinye Li, Rikhard Ranta and Padma Srivari carried out electrical characterization and data analysis. Andrea Padovani performed the defect analysis simulations and wrote the corresponding draft. Safdar Muhammad, Soumen Mazumder, Jahra Mariam and Xinye Li contributed to fabricating the devices. Sayani Majumdar contributed to supervision, resources, project administration, and funding acquisition. All authors discussed the results and commented on the manuscript.

Competing interests
The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41378-026-01429-4.

Received: 14 February 2026 Revised: 29 June 2026 Accepted: 28 July 2026
Published online: 04 September 2026

References

1. Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R. & Eleftheriou, E. Memory devices and applications for in-memory computing. *Nat. Nanotechnol.* **15**, 529–544 (2020).
2. Bhardwaj, K., Paasio, E. & Majumdar, S. Toward capacitive in-memory-computing: a device to systems level perspective on the future of artificial intelligence hardware. *Adv. Intell. Discov.* e202500143 (2025).
3. Yin, G. et al. Enabling lower-power charge-domain nonvolatile in-memory computing with ferroelectric FETs. *IEEE Trans. Circuits Syst. II: Express Briefs* **68**, 2262–2266 (2021).
4. Leroux, N. et al. Analog in-memory computing attention mechanism for fast and energy-efficient large language models. *Nat. Comput Sci.* **5**, 813–824 (2025).
5. Channamadhavuni, S., Thijssen, S., Jha, S. K. & Ewetz, R. Accelerating AI applications using analog in-memory computing: challenges and opportunities. In *Proc. Great Lakes Symposium on VLSI* 379–384. https://doi.org/10.1145/3453688.3461746. (ACM, 2021).
6. Yoo, J. et al. Recent research for HZO-based ferroelectric memory towards in-memory computing applications. *Electronics* **12**, 2297 (2023).
7. Ray, S. et al. Evaluation of a ferroelectric capacitor-based compute-in-memory framework. In *Proc. 31st International Conference on Electronics, Circuits and Systems (ICECS)* 1–4. https://doi.org/10.1109/ICECS61496.2024.10849145 (IEEE, 2024).
8. Majumdar, S. Back-End CMOS compatible and flexible ferroelectric memories for neuromorphic computing and adaptive sensing. *Adv. Intell. Syst.* **4**, 2100175 (2022).
9. Luo, Y.-C., Lu, A., Hur, J., Li, S. & Yu, S. Design and optimization of non-volatile capacitive crossbar array for in-memory computing. *IEEE Trans. Circuits Syst. II: Express Briefs* **69**, 784–788 (2022).
10. Hur, J. et al. Nonvolatile capacitive crossbar array for in-memory computing. *Adv. Intell. Syst.* **4**, 2100258 (2022).
11. Yu, E., Kumar, G. K., Saxena, U. & Roy, K. Ferroelectric capacitors and field-effect transistors as in-memory computing elements for machine learning workloads. *Sci. Rep.* **14**, 9426 (2024).

12. Zhou, Z. et al. Ferroelectric capacitive memories: devices, arrays, and applications. *Nano Converg.* **12**, 3 (2025).
13. Lederer, M., Lehninger, D., Ali, T. & Kämpfe, T. Review on the microstructure of ferroelectric hafnium oxides. *Phys. Rapid Res. Lett.* **16**, 2200168 (2022).
14. Zhang, D. et al. Unveiling ferroelectric HZO cryogenic performance (4–300 K): kinetic barrier engineering and underlying mechanism. *IEEE Trans. Electron Devices* **72**, 1788–1794 (2025).
15. Huang, F. et al. First observation of ultra-high polarization (108 μC/cm^2) in nanometer scaled high performance ferroelectric HZO capacitors with Mo electrodes. In *Proc. IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)* 1–2. https://doi.org/10.23919/VLSITechnologyandCir57934.2023.10185240 (IEEE, 2023),
16. Oh, S. et al. HfZrOx-based ferroelectric synapse device with 32 levels of conductance states for neuromorphic applications. *IEEE Electron Device Lett.* **38**, 732–735 (2017).
17. Bohuslavskyi, H., Grigoras, K., Ribeiro, M., Prunnila, M. & Majumdar, S. Ferroelectric Hf0.5Zr0.5O2 for analog memory and in-memory computing applications down to deep cryogenic temperatures. *Adv. Electron. Mater.* **10**, 2300879 (2024).
18. Hur, J. et al. Characterizing ferroelectric properties of Hf0.5 Zr0.5 O2 from deep-cryogenic temperature (4 K) to 400 K. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **7**, 168–174 (2021).
19. Singh Parihar, S. et al. Cryogenic hyperdimensional in-memory computing using ferroelectric TCAM. *IEEE J. Explorat. Solid-State Comput. Devices Circuits* **11**, 34–41 (2025).
20. Popovici, M. I. et al. High performance La-doped HZO based ferroelectric capacitors by interfacial engineering. In *Proc. International Electron Devices Meeting (IEDM)* 6.4.1–6.4.4 https://doi.org/10.1109/IEDM45625.2022.10019525 (2022).
21. Wang, S.-M., Liu, C.-R., Chen, Y.-T., Lee, S.-C. & Tang, Y.-T. Tunable defect engineering of Mo/TiON electrode in angstrom-laminated HfO2 /ZrO2 ferroelectric capacitors towards long endurance and high temperature retention. *Nanotechnology* **35**, 205704 (2024).
22. Li, X., Srivari, P., Honkanen, M., Salminen, T. & Majumdar, S. Record high polarization at 2V and imprint-free operation in superlattice HfO2-ZrO2 by proper tuning of ferro and antiferroelectricity. *Adv. Mater. Technol.* e02027 (2026).
23. Li, X., Srivari, P., Paasio, E. & Majumdar, S. Understanding fatigue and recovery mechanisms in Hf0.5 Zr0.5 O2 capacitors for designing high endurance ferroelectric memory and neuromorphic hardware. *Nanoscale* **17**, 6058–6071 (2025).
24. Lancaster, S. et al. Investigating charge trapping in ferroelectric thin films through transient measurements. *Front. Nanotechnol.* **4**, 939822 (2022).
25. Kazemi, A. et al. Achieving software-equivalent accuracy for hyperdimensional computing with ferroelectric-based in-memory computing. *Sci. Rep.* **12**, 19201 (2022).
26. Thomas, A., Dasgupta, S. & Rosing, T. A theoretical perspective on hyperdimensional computing. *J. Artif. Intell. Res.* **72**, 215–249 (2021).
27. Yun, S. et al. Hyperdimensional Computing for Robust and Efficient Unsupervised Learning. In *Proc. 57th Asilomar Conference on Signals, Systems, and Computers* 281–288 https://doi.org/10.1109/IEEECONF59524.2023.10476861 (2023).
28. Zhao, Z. et al. Engineering HZO by flat amorphous TiN with 0.3nm roughness achieving uniform c-axis alignment, record high breakdown field (10 nm HZO), and record final 2Pr of 56 μC/cm^2 with endurance > 4E12. In *Proc. IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits)* 1–2 https://doi.org/10.1109/VLSITechnologyandCir46783.2024.10631450 (2024).
29. Chen, G.-H. et al. Uniform and fatigue-free ferroelectric HZO with record EBD of 6.3MV/cm and record final 2Pr of 64 μCcm^2 at record 5E12 endurance using low lattice misfit (2.9%) beta-W. In *Proc. IEEE International Electron Devices Meeting (IEDM)* 1–4 https://doi.org/10.1109/IEDM50854.2024.10873405 (2024).
30. Hsain, H. A. et al. Role of oxygen source on buried interfaces in atomic-layer-deposited ferroelectric hafnia–zirconia thin films. *ACS Appl. Mater. Interfaces* **14**, 42232–42244 (2022).
31. Zhang, S. et al. Challenges and recent prospectives of 3D heterogeneous integration. *e-Prime - Adv. Electr. Eng., Electron. Energy* **2**, 100052 (2022).
32. Saha, J. K., Bukke, R. N., Mude, N. N. & Jang, J. Remarkable stability improvement of ZnO TFT with Al2O3 gate insulator by yttrium passivation with spray pyrolysis. *Nanomaterials* **10**, 976 (2020).
33. Strand, J., La Torraca, P., Padovani, A., Larcher, L. & Shluger, A. L. Dielectric breakdown in HfO2 dielectrics: using multiscale modeling to identify the critical physical processes involved in oxide degradation. *J. Appl. Phys.* **131**, 234501 (2022).
34. Ginestra Software. https://www.appliedmaterials.com/eu/en/semiconductor/ginestra-software.html.
35. Ron Cole, M. F. ISOLET. UCI machine learning repository https://doi.org/10.24432/C51G69 (1991).
36. Wasif, S. A. et al. Domain-specific hyperdimensional RISC-V processor for edge-AI training. *IEEE Trans. Circuits Syst. I: Regul. Pap.* **72**, 5825–5838 (2025).
37. Wei, W. et al. Deep insights into the failure mechanisms in field-cycled ferroelectric Hf0.5Zr0.5O2 thin film: TDDB characterizations and first-principles calculations. In *Proc. IEEE International Electron Devices Meeting (IEDM)* 39.6.1–39.6.4 https://doi.org/10.1109/IEDM13553.2020.9371932 (2020).
38. Vandelli, L. et al. A physical model of the temperature dependence of the current through SiO2HfO2 stacks. *IEEE Trans. Electron Devices* **58**, 2878–2887 (2011).
39. Muñoz Ramo, D., Gavartin, J. L., Shluger, A. L. & Bersuker, G. Spectroscopic properties of oxygen vacancies in monoclinic HfO_2 calculated with periodic and embedded cluster density functional theory. *Phys. Rev. B* **75**, 205336 (2007).
40. Perevalov, T. V. & Islamov, D. R. Atomic and electronic structure of oxygen polyvacancies in ZrO2. *Microelectron. Eng.* **178**, 275–278 (2017).
41. Gritsenko, V. A., Perevalov, T. V., Orlov, O. M. & Krasnikov, G. Ya. Nature of traps responsible for the memory effect in silicon nitride. *Appl. Phys. Lett.* **109**, 062904 (2016).
42. Pešić, M. et al. Variability sources and reliability of 3D — FeFETs. In *Proc. IEEE International Reliability Physics Symposium (IRPS)* 1–7. https://doi.org/10.1109/IRPS46558.2021.9405118 (2021).
43. Afroze, N. et al. Self-healing ferroelectric capacitors with ∼ 1000x endurance improvement at high temperatures (85–125°C). In *Proc. IEEE International Electron Devices Meeting (IEDM)* 1–4 https://doi.org/10.1109/IEDM50854.2024.10873595 (2024).
44. Zhou, H. et al. Application and benefits of target programming algorithms for ferroelectric HfO2 transistors. In *Proc. IEEE International Electron Devices Meeting (IEDM)* 18.6.1–18.6.) https://doi.org/10.1109/IEDM13553.2020.9371975 (2020).
45. Pesic, M., Padovani, A., Slcsazeck, S., Mikolajick, T. & Larcher, L. Deconvoluting charge trapping and nucleation interplay in FeFETs: Kinetics and Reliability. In *Proc. IEEE International Electron Devices Meeting (IEDM)* 25.1.1–25.1.4 https://doi.org/10.1109/IEDM.2018.8614492 (2018).
46. Srivari, P., Paasio, E., Anu, A. & Majumdar, S. Effect of thermal engineering on the reliability of analog states in ferroelectric Hf0.5Zr0.5O2. In *Proc. IEEE International Integrated Reliability Workshop (IIRW)* South Lake Tahoe, CA, USA, 1–5 https://doi.org/10.1109/IIRW66154.2025.11481754 (IEEE, 2025).
47. Paasio, E., Ranta, R. & Majumdar, S. A physics-based compact model for ferroelectric capacitors operating down to deep cryogenic temperatures for applications in analog memory and neuromorphic architectures. *Adv. Electon. Mater.* **11**, 2400840 (2025).
48. Larcher, L., Padovani, A., Puglisi, F. M. & Pavan, P. Extracting atomic defect properties from leakage current temperature dependence. *IEEE Trans. Electron Devices* **65**, 5475–5480 (2018).
49. Sun, X. et al. PCM-based analog compute-in-memory: impact of device non-idealities on inference accuracy. *IEEE Trans. Electron Devices* **68**, 5585–5591 (2021).
50. Cai, H. et al. Proposal of analog in-memory computing with magnified tunnel magnetoresistance ratio and universal STT-MRAM cell. *IEEE Trans. Circuits Syst. I: Regul. Pap.* **69**, 1519–1531 (2022).
51. Cassuto, Y., Kvatinsky, S. & Yaakobi, E.Sneak-path constraints in memristor crossbar arrays, In *Proc. IEEE International Symposium on Information Theory* 156–160 (IEEE, 2013).
52. Zhou, J., Kim, K.-H. & Lu, W. Crossbar RRAM arrays: selector device requirements during read operation. *IEEE Trans. Electron Devices* **61**, 1369–1376 (2014).
53. Khwa, W.S. et al., MRAM design–technology–system co-optimization for artificial intelligence edge devices, In *Proc. IEEE International Electron Devices Meeting (IEDM)*, 1–4 (IEEE, 2024).
54. Roldán, J. B. et al. Variability in resistive memories. *Adv. Intell. Syst.* **5**, 2200338 (2023).
55. Burr, G. W. et al. Phase change memory technology. *J. Vac. Sci. Technol. B.* **28**, 223–262 (2010).
56. Zhou, Z. et al. Inversion-type ferroelectric capacitive memory and its 1-Kbit crossbar array. *IEEE Trans. Electron Devices* **70**, 1641–1647 (2023).

57. Chen, C. et al. Understanding the impact of programming pulses and electrode materials on the endurance properties of scaled Ta2O5 RRAM cells. In *Proc. IEEE International Electron Devices Meeting* (IEEE, 2014).
58. Hellenbrand, M., Teck, I. & MacManus-Driscoll, J. L. Progress of emerging non-volatile memory technologies in industry. *MRS Commun.* **14**, 1–14 (2024).
59. Wouters, D. J., Waser, R. & Wuttig, M. Phase-change and redox-based resistive switching memories. *Proc. IEEE* **103**, 1274–1288 (2015).
60. Jeon, K. et al. Self-rectifying resistive memory in passive crossbar arrays. *Nat. Commun.* **12**, 2968 (2021).
61. Cheng, H. Y., Carta, F., Chien, W. C., Lung, H. L. & BrightSky, M. J. 3D cross-point phase-change memory for storage-class memory. *J. Phys. D: Appl. Phys.* **52**, 473002 (2019).
62. Lin, H. & Jiang, Y. Optimization of the 3D multi-level SOT-MRAMs. *AIP Adv.* **14**, 025002 (2024).
63. Waser, R. Resistive non-volatile memory devices. *Microelectron. Eng.* **86**, 1925–1928 (2009).
64. Le Gallo, M. & Sebastian, A. An overview of phase-change memory device physics. *J. Phys. D: Appl. Phys.* **53**, 213002 (2020).
65. Ielmini, D. Brain-inspired computing with resistive switching memory (RRAM): devices, synapses and neural networks. *Microelectron. Eng.* **190**, 44–53 (2018).
66. Wan, W. et al. A compute-in-memory chip based on resistive random-access memory. *Nature* **608**, 504–512 (2022).
67. Soliman, T. et al. First demonstration of in-memory computing crossbar using multi-level cell FeFET. *Nat. Commun.* **14**, 6348. https://doi.org/10.1038/s41467-023-42110-y (2023).