Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version XX Month, XXXX.

Digital Object Identifier 10.1109/XXXX.2022.1234567

Analysis of FeFET-based Analog Synapse

Sayma Nowshin Chowdhury[1†], IEEE Student Member, Alex L. Mazzoni[2†], Xiaohang Zhang [2], Andreu L. Glasmann [2], Sina Najmaei[2], and Sahil Shah [1], Senior Member, IEEE

[1]University of Maryland, College Park, MD USA
[2]Army Research Laboratory, Adelphi, MD, USA

Corresponding author: Sina Najmaei (email: sina.najmaei.civ@army.mil), Sahil Shah (email: sshah389@umd.edu).

[†]**These authors contributed equally to this work.**

This work was partly supported by National Science Foundation award #2210804. Sayma Nowshin Chowdhury is funded via Army Research Laboratory under Cooperative Agreement Number W911NF-24-2-0048 starting January 2024.

ABSTRACT

INDEX TERMS Ferroelectric Field-Effect Transistors, Verilog-A, Analog Synapses

I. FeFET as an Analog Synapse

The exponential growth in data-driven applications, especially in artificial intelligence (AI) and machine learning, has underscored the limitations of traditional von Neumann architectures. These architectures suffer from a significant memory and computation bottleneck, particularly when dealing with deep learning workloads that involve large volumes of data and extensive synaptic weight storage. Neuromorphic architectures address this bottleneck by physically co-locating memory and computation. In neuromorphic computing, analog synapses store weights of neural networks and effectively function as non-volatile, variable conductance states to perform in-memory computation. Conventional memory elements, such as static random-access memory (SRAM) and dynamic random-access memory (DRAM), are unsuitable for analog synaptic functions, especially in edge applications, due to their binary states, large footprint, and energy inefficiency in maintaining data over time. Emerging technologies such as resistive RAM (ReRAM) and phase change memory (PCM) have demonstrated potential in analog memory applications but face challenges related to non-linearity, variability, limited endurance under high-speed operation, device architecture, and integration challenges [1]. The ferroelectric field-effect transistor (FeFET) has emerged as a promising alternative with low-power operation, high speed, and compatibility with current complementary metal-oxide-semiconductor (CMOS) front-end-of-line (FEOL) processes, making it highly suitable for applications in energy-constrained environments such as IoT and edge-AI devices. As a three-terminal device, the FeFET offers enhanced control over channel conductance, which is critical for analog weight modulation. By leveraging its non-volatile characteristics and ability to achieve multi-level storage, the FeFET stands out as a robust candidate for analog synaptic applications. The unique characteristics of the FeFET enable its application as an analog synapse in neuromorphic computing systems, providing a viable path for overcoming data movement limitations and accelerating machine learning processes.

This paper examines the properties of FeFETs when used as an analog synaptic device for in-memory computing and its implications for neuromorphic systems. To assess the characteristics of these ferroelectric devices, we perform extensive electrical characterization of GlobalFoundries' FeFETs based on ferroelectric hafnia and their 28nm SLPe process. We will cover the FeFET's structural, electrical, and operational attributes that make the device particularly suitable for tasks like weight storage, analog multiply-accumulate (MAC) operations, and real-time learning, especially in edge computing contexts where low power and high-speed processing are essential. While there has been significant progress in the development of FeFET technology [2]–[4], there are still existing gaps, especially in the analog, in-memory computing realm, that this research addresses. Specifically, data is often presented for binary state retention but not as often for intermediate states. Here we show 3-bit retention data for both n-type (NMOS) and p-type (PMOS) FeFETs at room temperature and 125 °C. To better explore the control of these multilevel conductances, we go beyond normal potentiation and depression and investigate in detail

the impact of FeFET size, again for both NMOS and PMOS devices. Furthermore, papers on device-level testing can neglect the link to practical circuit- or system-level integration considerations, with research and reporting occurring in a siloed manner. Here, whenever possible, we try to tie device performance to chip design considerations such as the complexity/size/power/cost of peripheral circuitry. We also introduce and discuss multiple fundamental tradeoffs that we see as design decisions, which are highly application-specific.

II. FeFET Functional Principles

A. Switching Dynamics and Operation

FeFET devices replace or augment the normal dielectric found in the gate stack of a FET with a ferroelectric layer, commonly made of solid solutions of HfO_2 and interstitial such as Si or Zr. Remanent ferroelectric polarization in the gate stack enables distinct and reconfigurable, non-volatile threshold voltages (V_{th}) in the transistor channel. This ferroelectric polarization can be switched by controlling the electric field across the ferroelectric layer, typically done via voltage pulses to the gate terminal.

The dynamics of switching involves nucleation and domain wall propagation within the ferroelectric layer. Initially, under a strong enough electric field, small regions (nuclei) of reversed polarization are formed. These nuclei grow as the domain walls move, eventually reversing the entire polarization if the voltage is maintained or increased. This domain-level behavior underpins the macroscopic switching characteristics observed in FeFET devices [5].

The switching mechanism is inherently fast and energy-efficient, making single-bit FeFETs suitable for traditional non-volatile memory applications. Notably, sub-nanosecond switching behavior has been demonstrated in HfO_2-based FeFETs, where the polarization reversal time was observed to span 11 orders of magnitude for only a single order change in pulse voltage. This steep time-voltage relationship is well captured by classical nucleation theory, confirming the voltage-sensitive kinetics of domain nucleation and switching in ferroelectric materials [6].

1) Single-Bit Operation (Traditional Memory)

Single-bit operation in FeFETs represents the most fundamental mode of device functionality, wherein the device toggles between two polarization states—corresponding to binary logic levels. In this mode, switching dynamics are governed by the complete reversal of ferroelectric polarization when the applied gate voltage exceeds the coercive field (E_c) of the material. This process leads to two distinct threshold voltages, enabling the binary storage of '0' and '1' as illustrated in Figure 1(a).

The two-state behavior arises from the homogeneous switching of all ferroelectric domains once the coercive field is surpassed. However, the precision and endurance of single-bit FeFETs depend heavily on material quality, domain uniformity, and interface stability between the ferroelectric and semiconductor layers, and the device architecture (e.g., MFIS (Metal-Ferroelectric–Insulator–Semiconductor) vs. MFMIS (Metal–Ferroelectric–Metal–Insulator–Semiconductor) structures) [3], [7], [8].

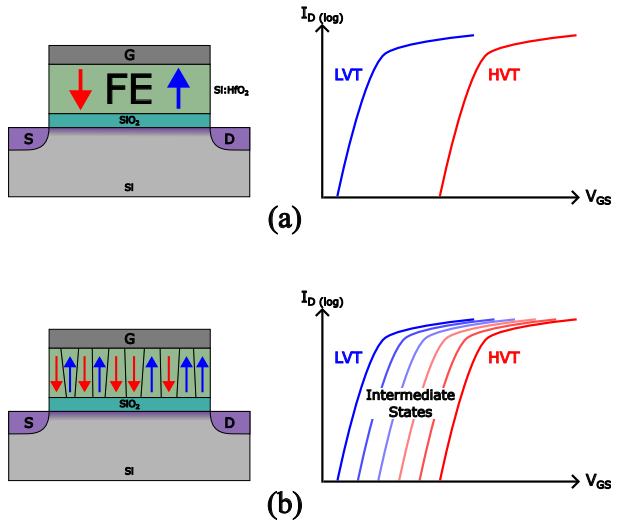

FIGURE 1. Example FeFET operation for an n-type channel. (a) Single-Bit Operation: Ferroelectric polarization is predominantly all in either the up or down state. This leads to two distinct states, (red) the high threshold voltage (HVT) and (blue) low threshold voltage (LVT) states. (b) Multi-Bit Operation: By applying voltage pulses with amplitudes and widths lower than a critical value, only a subset of the ferroelectric domains will switch, leading to intermediate states between the HVT (red) and LVT (blue) states.

2) Multi-Bit Operation (Analog Memory)

Multistate performance in FeFETs significantly enhances their utility in neuromorphic and memory-intensive applications by enabling the storage of more than two conductance levels per cell. Unlike binary switching, multistate behavior leverages the partial switching of ferroelectric domains due to the heterogeneous coercive field distribution across the ferroelectric layer.

In an idealized mono-domain system, the coercive field (E_c)—the electric field required to switch the polarization direction—is sharply defined and uniform across the material. However, practical ferroelectric layers, especially polycrystalline HfO_2 solid solutions, exhibit significant heterogeneity due to variation in grain morphology, crystallographic orientation, interstitial distribution, and internal mechanical stress. These microstructural non-uniformities lead to a broad distribution of coercive fields among the domains in the ferroelectric layer [9].

Domains with lower E_c switch at lower voltages, while others require higher fields. This physical phenomenon allows for gradual and selective switching of ferroelectric domains by carefully tuning the amplitude and duration

of voltage pulses. As a result, the material can be partially polarized, and the overall polarization state becomes a weighted sum of the switched and unswitched domains. This intermediate polarization gives rise to multiple discrete threshold voltage levels in FeFETs, forming the basis for multi-bit operation as illustrated in Figure 1(b).

This behavior is quantitatively described using physical models such as the Preisach model, which has long been employed to simulate polarization switching and hysteresis in ferroelectric capacitors and memory devices [10]. The model treats the ferroelectric material as a superposition of bistable domains, each with its own switching threshold, and it can accurately replicate the hysteresis behavior observed experimentally in multistate memory systems. In experimentally fabricated FeFETs using solid solutions of HfO_2, the heterogeneous distribution of coercive fields enables multilevel threshold voltages due to selective domain activation. Mulaosmanovic et al. demonstrated that precise voltage pulse control can induce stable intermediate polarization states, which correspond to distinct threshold voltages in the device's transfer characteristics. Each of these voltage states translates into a unique drain current level under fixed gate bias conditions, allowing analog or multibit conductance representation [11].

Device geometry and scaling significantly affect multistate performance. Larger FeFET devices typically contain a greater number of ferroelectric grains, averaging out the effects of domain variability and yielding smoother and more predictable analog behavior [12]. In contrast, nanoscale devices contain fewer grains and are more susceptible to stochastic switching behavior due to the pronounced influence of individual domain variability [13].

Further advancements have been achieved through material and device engineering. Zeng et al. showed that engineering the electric field gradient across the device stack can enhance multistate capability. By spatially modulating the coercive field distribution within the ferroelectric layer, they achieved finer control over domain switching, resulting in increased multistate resolution [14]. Experimental demonstration was performed on a 500 nm channel length, with simulation demonstrating its use down to 20 nm. This technique underscores the importance of internal field engineering for fine-tuning device behavior. In a similar vein, the "MirrorBit" FeFET concept has been demonstrated on a 240 nm x 240 nm device as a way to convert 1-bit FeFETs to 2-bit FeFETs through the control of drain and source terminal voltages [15].

B. Synaptic Functionality in FeFETs: Potentiation and Depression

The functional role of an analog synapse in neuromorphic computing largely depends on its ability to emulate the behavior of biological synapses by adjusting synaptic weights in response to external stimuli. In FeFETs, this adjustment process is achieved through potentiation and depression mechanisms, which are responsible for increasing or decreasing the conductance of the device, respectively. These processes enable FeFET devices to act as analog synapses, simulating the strengthening and weakening of synaptic connections in artificial neural networks.

In contrast to charge-based memory technologies such as DRAM or FeRAM, FeFETs do not rely on a separate storage capacitor. Instead, the polarization state of the gate stack controls the channel's electrical properties through electrostatic modulation of the potential barrier, making the read operation non-destructive. This is particularly advantageous in neuromorphic and in-memory computing systems, where frequent read operations are required during inference and training cycles [16], [17].

1) Potentiation (Increasing Conductance)

Potentiation in FeFETs involves the incremental increase of channel conductance by systematically shifting the device's threshold voltage (V_{th}) through the alignment of ferroelectric dipoles. This is achieved by applying a series of carefully controlled voltage pulses (positive for NMOS, negative for PMOS) to the gate terminal. Each pulse incrementally aligns more dipoles in the ferroelectric layer in the same direction as the field, resulting in increased remanent polarization. This shift in V_{th} leads to a higher channel current for a given gate voltage, simulating synaptic strengthening in artificial neural networks [1].

Notably, the degree of potentiation depends on several factors, including pulse amplitude, duration, and frequency. Smaller, repeated pulses result in finer resolution of conductance states, which is essential for enabling multi-level analog weight representation. This granularity allows FeFET-based synapses to encode a spectrum of weight values, rather than binary states alone. The controllability and retention of the potentiated state are further influenced by material selection (e.g., doped HfO_2), device geometry, and the ferroelectric domain dynamics [18].

2) Depression (Decreasing Conductance)

Conversely, depression in FeFETs is the mechanism by which the conductance is decreased—representing the weakening or removal of synaptic connections in neural architectures. This is implemented by applying a series of voltage pulses (negative for NMOS, positive for PMOS) that change V_{th} and reduce the drain current, reaching a lower conductance state [1].

The depression process, like potentiation, is ideally gradual and controllable. The symmetry between potentiation and depression ensures that learning algorithms can both enhance and suppress weights effectively, enabling backpropagation-based learning and stability during training. However, achieving symmetric behavior in practice requires precise calibration of pulse schemes and careful engineering

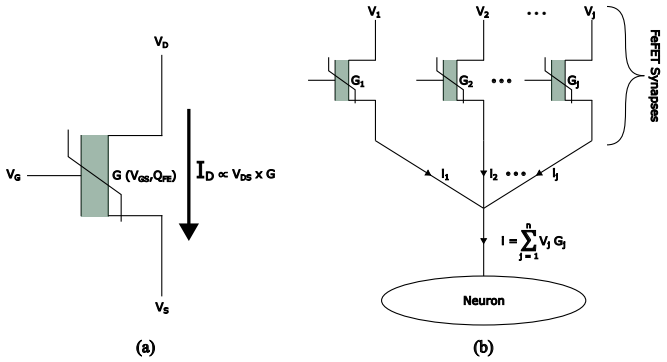

FIGURE 2. (a) Synaptic weights are encoded as remanent polarization or charge in the ferroelectric, Q_{FE}. Conductance, G, of each FeFET is a function of Q_{FE} and the chosen gate read voltage, V_{GS}. (b) FeFET synapses with a shared node intrinsically perform the multiply-accumulate (MAC) function, a core function in artificial neural networks.

of ferroelectric layers to avoid hysteresis asymmetry [19]. Importantly, this reversibility is key to the dynamic plasticity of artificial synapses, allowing for weight updates across multiple epochs without device degradation.

Multi-Level Synaptic Weights:

The ability to achieve multiple conductance states makes FeFETs highly suitable for implementing synaptic weights in spiking neural networks and deep learning accelerators. The analog modulation of weights is facilitated by the progressive and partial polarization switching of ferroelectric domains as described above. These mechanisms support symmetric weight updates across multiple levels, critical for neuromorphic learning algorithms such as backpropagation and Hebbian learning. The progressive nature of switching also helps mitigate abrupt transitions and allows for finer control over network dynamics, resulting in more efficient and accurate neural computation [20].

However, the conductance (weight) of a FeFET is not solely a function of the ferroelectric polarization, Q_{FE}. The dynamic range and linearity of the conductances (weights) is highly dependent on the choice of the gate voltage, V_{GS}, during the read process (Figure 2(a)). For example, at V_{GS} values closer to subthreshold the dynamic range gets larger while the linearity deteriorates, whereas for V_{GS} closer to the linear regime the dynamic range shrinks and linearity improves. As long as the gate voltage is low enough to not disturb the ferroelectric polarization, the choice of V_{GS} (also channel doping, gate work function, etc.) becomes a powerful tool to optimize network performance.

In neuromorphic and in-memory computing architectures, FeFETs are often arranged in crossbar arrays where each FeFET functions as a programmable synapse, and multiple synapses are connected to an artificial neuron circuit . This arrangement enables analog computation of the multiply-accumulate function as shown in Figure 2(b). This combination of multiple synapses that lead to an artificial neuron is mappable into a crossbar array architecture to accelerate AI workloads.

III. FeFET Synapse Programming Schemes

As mentioned earlier, pulsing schemes can generally be categorized as controlling pulse amplitude, pulse width, and/or pulse count. The general comparison of controlling pulse height, width, or applying identical pulses has been demonstrated before, with variable pulse amplitude shown to provide the best control for multi-state operation [21]. Other choices include whether or not to have a RESET pulse every time to have a more established starting point, or when it comes to implementing these programming schemes on a system or chip, there is also the choice of feedback or control to achieve the desired conductance state.

Accurate evaluation of FeFET device performance necessitates carefully designed electrical measurement protocols. A widely adopted methodology employs a pulse sequence comprising **RESET–DELAY–SET–DELAY–READ** operations [22]–[25]. The RESET pulse is followed by a delay interval to allow for transient effects—such as charge redistribution, interface trap relaxation, and incomplete ferroelectric stabilization—to subside. A second delay is again introduced after the SET pulse to ensure that polarization dynamics and charge detrapping are complete prior to the DC READ sweep, which measures either the output current or the threshold voltage.

These delay periods are not arbitrary. Without sufficient delay, the measurement may include transient contributions from unstable charge states or dynamic polarization relaxation, leading to erroneous conclusions regarding the device's stable memory state. Beyer *et al.* illustrated this issue through the *sloshing bathtub model*, wherein rapid switching induces electron accumulation at the ferroelectric/interlayer (FE/IL) interface, accelerating charge trapping into deeper energy states and resulting in premature V_{th} walkout and endurance degradation [22]. This is particularly problematic in *n*-type FeFETs, where the conductance level stabilizes well after the switching event.

The necessity of proper delay intervals was further corroborated by Mulaosmanovic *et al.*, who demonstrated that omitting delays between SET/RESET and READ operations significantly misrepresents the effective memory window, especially in scaled FeFETs [26]. Mulaosmanovic *et al.* expanded on this work by proposing optimized timing protocols to isolate intrinsic ferroelectric switching from extrinsic parasitic effects. These refined pulse schemes are essential for accurate reliability benchmarking and for minimizing read disturbance in multilevel programming scenarios [27].

A. 20 mV Incremental Pulse Scheme

In this study, an incremental amplitude pulse scheme following the RESET-DELAY-SET-DELAY-READ method was applied to FeFETs of various sizes fabricated in the GlobalFoundries 28 nm node. In this method, programming and erasing pulses were applied with progressively increasing amplitude steps (e.g., 2 V to 5.0 V in 20 mV increments), while the duration remained fixed. The choice of a 20 mV

step size was motivated by a 5 V range, subdivided by an 8-bit digital-to-analog converter (DAC), which is roughly 20 mV.

As described previously, 5 s delay intervals followed each pulse to ensure that the observed V_{th} reflected a more stabilized state of the device. RESET pulse widths were 1 µs, while the SET pulse widths were 500 ns. Specific pulse train schematics for both potentiation and depression are shown in Figures 3 and 4, respectively, with experimental data for both NMOS and PMOS devices below the pulse schemes. We stress that this testing procedure includes a RESET before each measurement or V_{GS} sweep; therefore, *it tests not only the number of available states but also the repeatability or stochasticity of setting a state*.

FeFETs were tested on a FormFactor SUMMIT200 probe station and measured using a Keysight B1500A semiconductor device analyzer. Gate pulses were applied via the B1525A high-voltage source/pulse generator unit (HV-SPGU), while voltage biasing and current measurements were conducted using B1511B source-measure units (SMUs). The drain, source, and bulk terminals were grounded during gate pulsing. To enable pulse application and measurement on the same physical probe, the SPGU output and gate SMU were connected using a Keithley 4205 remote bias-tee. This configuration ensured precise control of pulse delivery and synchronization with transient relaxation periods.

While this pulsing scheme was chosen to enable finer control over the polarization process for device characterization, it should be noted that in a real embedded system these pulses need to be created on-chip which can increase the complexity (size, cost, power, etc.) of the peripheral circuitry.

Overall, the incremental amplitude scheme offers significantly enhanced weight resolution compared to the fixed pulse method, making it better suited for implementing analog synapses in neuromorphic systems where precise conductance tuning is required.

IV. FeFET Synapse Performance Metrics

We mainly focus on two aspects of these FeFET analog synapses, their intermediate states and the retention of those states. However, other important performance metrics will be introduced and discussed, such as the linearity of potentiation and depression behavior, stochasticity in setting a state, chip space occupied, and energy consumption.

A. Intermediate States

A balanced understanding of both NMOS and PMOS FeFETs is important for designing analog synapses in neuromorphic computing systems. While the fundamental switching mechanism—reliant on polarization reversal in the ferroelectric layer—is similar in both device types, differences in band alignment, carrier mobility, interface behavior, and trap dynamics influence their analog characteristics. Experimental measurements conducted on GlobalFoundries 28nm FeFET devices show that both NMOS and PMOS variants are capable of stable multistate switching, suitable for analog computing applications, though with distinct performance profiles. The main difference between the NMOS and PMOS FeFETs we tested is the magnitude of the V_{th} offset from 0 V_{GS}, leading to slightly asymmetric biasing and response between the two types of devices.

An ideal analog synapse has access to a wide continuum of conductance states. It is clear from Figures 3 and 4 that large-area (7 µm x 10 µm) FeFETs come close to this continuum, with most of the space bounded by the LVT and HVT filled with intermediate states. However, for smaller FeFETs, the available intermediate states become limited as illustrated in Figure 5.

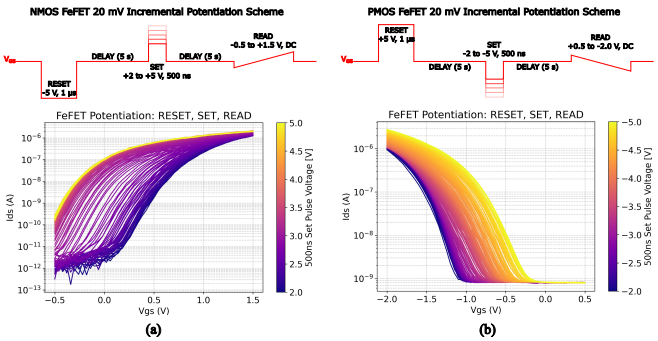

FIGURE 3. Potentiation characteristics measured for incremental program pulses of (a) 2 V to 5 V (500 ns pulse width) on a 7 µm x 10 µm NMOS FeFET device (b) -2 V to -5 V (500 ns pulse width) on a 7 µm x 10 µm PMOS FeFET device

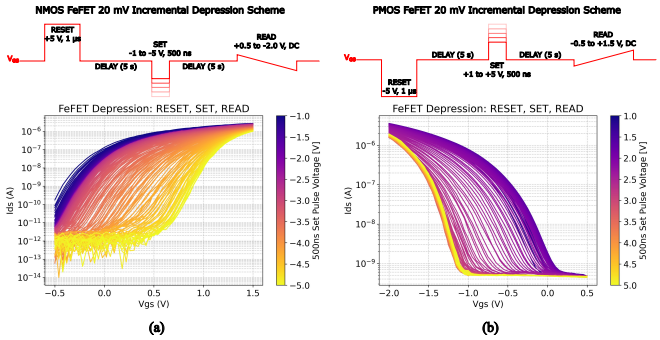

FIGURE 4. Depression characteristics measured for incremental program pulses of (a) 2 V to 5 V (500 ns pulse width) on a 7 µm x 10 µm NMOS FeFET device (b) -2 V to -5 V (500 ns pulse width) on a 7 µm x 10 µm PMOS FeFET device

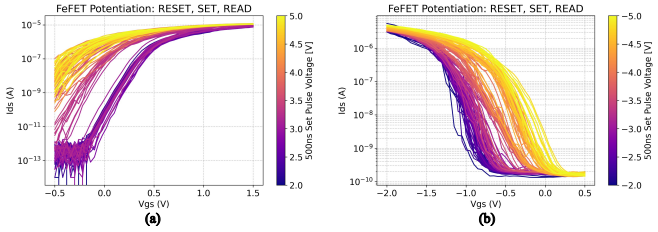

FIGURE 5. Potentiation characteristics of the smallest measured NMOS FeFET (a), 80 nm x 34 nm, and the smallest measured PMOS FeFET (b), 200 nm x 200 nm.

TABLE 1. Estimate for Number of Grains in Each Device Size

One Grain Size* [nm]	One Grain Area [nm^2]	Device Width [nm]	Device Length [nm]	Total Grains Upper Bound	Total Grains Lower Bound
20	400	7000	10000	175000	7000
100	10000	1000	1000	2500	100
		500	500	625	25
		200	200	100	4
		80	34	6.8	0.272

*Lateral grain size typically 20-30 nm, up to 100 nm.

1) Impact of Device Size

A larger FeFET contains more ferroelectric domains than a smaller FeFET. This leads to a larger distribution of coercive fields, which enables not only more states — it also enables a more deterministic programming of the intermediate states. One can glean this information from our testing results by examining the color gradient in Figures 3 through 5. If the color gradient in the $I_D - V_{GS}$ curves closely matches the gradient in the colorbar scale, that means that the FeFET potentiation or depression (with a RESET pulse between every curve) is behaving in a repeatable manner. In a study by Alessandri *et al.*, [28] they demonstrated that the stochastic nucleation limited switching dynamics of a ferroelectric capacitor shows substantial variability in devices with 100 grains but is virtually indistinguishable from the mathematical average at 5000 grains.

The lateral grain size in these ferroelectric hafnia thin films are typically around 20-30 nm [29]. For the device sizes tested in this study, this leads to the estimate of grains contained in each device shown in Table 1.

The trend in our measurements agrees well with that of Alessandri *et al.* The largest devices tested (7 µm x 10 µm) enable a more smooth and deterministic setting of states; this agrees with our estimate that these devices include many thousands of ferroelectric grains. However, with smaller FeFETs, there is clear variability in the setting of conductance states. This can be viewed more directly by taking vertical slices in the 20 mV incremental pulsing data at specific V_{GS} values. Figure 6 plots an example slice of normalized potentiation data for both NMOS and PMOS FeFETs of different sizes. Once the FeFET size reaches 7 µm x 10 µm (black data points), the potentiation becomes a smooth, monotonically increasing function of the SET pulse magnitude, indicating a high level of repeatability or determinism in setting a conductance level. Smaller devices, on the other hand, show a more variable relationship to the magnitude of the SET pulse, meaning that there is inconsistent programming of the conductance level.

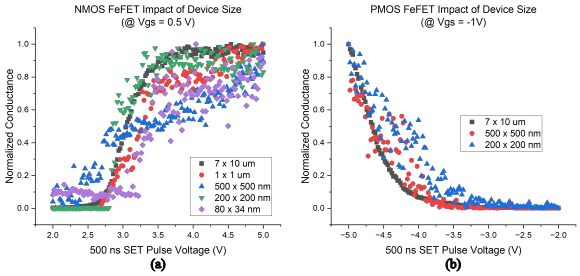

FIGURE 6. Effect of FeFET Device Size on Normalized Conductance and State Density for both NMOS (a) and PMOS (b) devices.

B. Retention of States

A perfect synapse would store its programmed weight without any degradation over time or from a change in operating conditions, such as temperature. In reality, non-volatile memory typically suffers some data loss, often specified at an elevated temperature and a specific time frame. To characterize retention characteristics, devices are typically measured for 10^3 to 10^5 s at an elevated temperature and then extrapolated to a 10-year retention figure of merit.

The largest devices (7 µm x 10 µm) were chosen for retention testing as their state setting behaves the most deterministically. Following the same RESET-DELAY-SET-DELAY-READ programming scheme, conductance levels were set and then read at a specific V_{GS} for 1000 s. Retention data was measured for at least 8 states for NMOS and PMOS devices at both room temperature and 125 °C (see Figure 7). The retention data fit well to a power law description, with

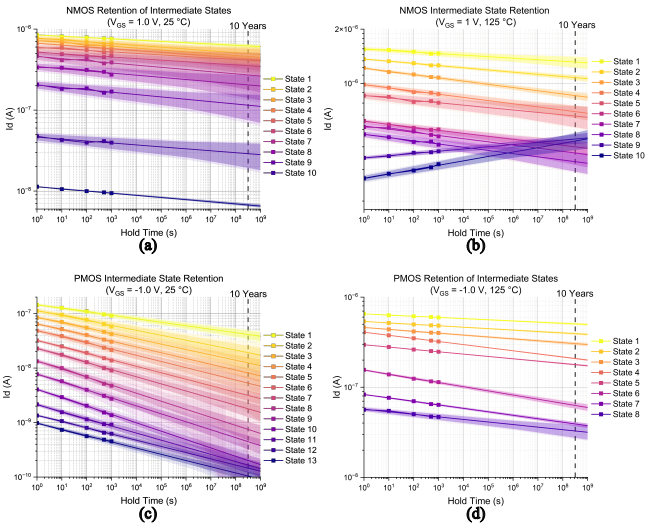

FIGURE 7. Retention of intermediate states in NMOS devices at room temperature (a) and at 125 ℃ (b), PMOS devices at room temperature (c) and 125 ℃ (d). Each plot shows the fitted line (solid) as well as the 95% confidence band (darker shade) and 95% prediction band (lighter shade).

values for A being roughly the starting current at 0 s and ranges for b as listed in table 2.

$$I(t) = At^b \tag{1}$$

TABLE 2. Extracted Power Law Constants

Device Type	b (Lower Bound)	b (Upper Bound)
NMOS (25 C)	-0.016	-0.029
NMOS (125 C)	-0.0085	0.025
PMOS (25 C)	-0.06	-0.15
PMOS (125 C)	-0.006	-0.046

The impact of charge trapping on FeFET behavior, particularly in NMOS devices, has been discussed in earlier sections. In those discussions, we noted that electron trapping near the interface between the ferroelectric HfO_2 layer and the underlying semiconductor can lead to time-dependent threshold voltage shifts, particularly when insufficient delay is provided after switching pulses. This dynamic contributes to transient instability and can obscure the true remnant polarization state, especially in scaled devices or when operating at high speeds [2], [30].

In this section, we extend the analysis by examining trap-related behavior in PMOS FeFETs, which was not previously addressed. Unlike NMOS devices, PMOS FeFETs predominantly involve hole transport, and thus exhibit different trapping and detrapping characteristics. Experimental observations show that PMOS devices demonstrate better retention performance over longer periods. This is attributed to the slower detrapping kinetics of holes, which occupy deeper energetic states in the ferroelectric material compared to electrons. These deeper traps are less thermally activated at room temperature, resulting in reduced spontaneous charge loss and more stable threshold voltages [30], [31].

V. Discussion

A. Fundamental Design Trade-offs

1) FeFET Size

There are two main strategies by which "analog" FeFET synapses can be achieved. Large-area FeFETs are the only way to provide true >3 bit analog behavior, whereas small FeFETs will likely rely on the combination of individual 1-2 bit FeFETs to reach the desired bit precision. While the large FeFETs enable more intermediate states and have more deterministic access to those states, it comes at the price of area and programming energy. The energy used to program an FeFET is directly proportional to the ferroelectric switching area, coercive field (E_C), and the remanent polarization (P_R). The area ratio of the largest FeFET tested to the smallest is >25,000 which directly reflects the switching energy ratio. To reap the efficiency of small FeFETs, there must also be a higher overhead of peripheral circuitry to address the higher stochasticity and routing required for 1- or 2-bit FeFETs. One strategy to decrease the variability and read energy of smaller FeFETs is the use of a series resistor to limit V_{th} shifts and lower the overall current [32].

A general comparison of large and small FeFETs to traditional devices like SRAM can be found in table 3.

B. Retention Analysis

This retention advantage is particularly relevant in neuromorphic applications where the stability of stored analog weights over time is critical. The physical mechanisms underlying this behavior suggest that PMOS FeFETs may be better suited for long-term analog memory functions, while NMOS FeFETs remain favorable for faster, high-frequency switching applications.

The physical dimensions of FeFET devices strongly influence the number and stability of achievable conductance states. Larger devices tend to include more ferroelectric grains in the active channel. Since each grain may have a slightly different coercive field due to structural variations, devices with larger lateral dimensions exhibit a broader distribution of polarization switching thresholds. This enables a finer granularity of intermediate states and more continuous conductance modulation [26].

Both NMOS and PMOS devices in this study showed an increase in the number of accessible states as the lateral dimensions increased from 200 nm × 200 nm to 7 µm × 10 µm. The enhanced multistate resolution in larger devices is attributed to statistical averaging over a larger number of grains, which minimizes the variability seen in smaller devices. These effects are consistent across device types, suggesting that geometry plays a central role in determining synaptic resolution, regardless of whether the device is NMOS or PMOS.

Smaller devices, in contrast, are more susceptible to stochastic switching behavior. With fewer active domains, the effect of individual grain switching becomes more pronounced, leading to variability in the conductance response. This imposes a limitation on how aggressively device dimensions can be scaled while still maintaining reliable multistate operation.

The comparative analysis shows that both NMOS and PMOS FeFETs can serve as effective analog synaptic elements, with distinct advantages depending on the intended use case. NMOS devices generally benefit from faster switching and stronger drive currents, which can be useful in high-speed or low-latency applications. PMOS devices, meanwhile, offer improved data retention and are less sensitive to read disturbances, which may be important in systems requiring long-term analog state storage.

The retention benefit observed in PMOS FeFETs can be understood from a physical perspective as arising from the deeper energy levels of hole traps and the reduced

TABLE 3. Comparison of Large and Small FeFET metrics

Metric	Large FeFET (7 µm x 10 µm)	Small FeFET (80 nm x 34 nm)
Number of States	>3-4 bit	1-2 bit
Intermediate State Programming Behavior	Deterministic	Stochastic
Area (μm^2)	70	0.00272
Programming Energy (J)	10-100 pJ	1-10 fJ
METRIC OF INTEREST	INSERT	HERE
METRIC OF INTEREST	INSERT	HERE
METRIC OF INTEREST	INSERT	HERE

TABLE 4. Comparison of FeFET and SRAM-Based Analog Synapses

Technology	**Node**	**Foundry**	**Footprint** (μm^2)	**# States**	**Bit Density (bits/μm^2)**	**Notes**
FeFET	28 nm	GF		8–16		Multi-bit analog storage enabled by incremental pulsing
6T-SRAM (Standard)	10 nm		0.0312	2 (binary)		Minimal area; single-bit storage with standard read/write
6T-SRAM (Standard)	5 nm	TSMC	0.021	2 (binary)		Scaled bitcell size; density improves, but storage remains binary without circuit-level enhancements [33].
SRAM-CIM	28 nm		8.71	3–13 (analog levels)	0.34	Local computing cell (LCC) shared across 8 SRAM cells; dual-mode computation supports 2b×1b MAC operations.

TABLE 5. Retention Time for Different Voltages at 25 C

Voltage (V)	I_0 **(A)**	$I_{\mathbf{10yr}}$ **(A)**	**% Drop**
2.30	4.75×10^{-8}	2.65×10^{-8}	44.30
2.50	2.07×10^{-7}	9.81×10^{-8}	52.52
2.60	3.37×10^{-7}	1.78×10^{-7}	47.24
2.70	4.66×10^{-7}	2.39×10^{-7}	48.67
2.80	5.24×10^{-7}	3.34×10^{-7}	36.27
2.90	5.83×10^{-7}	3.90×10^{-7}	33.09
3.10	6.68×10^{-7}	3.94×10^{-7}	41.03
3.30	7.19×10^{-7}	4.22×10^{-7}	41.31
3.50	7.65×10^{-7}	4.56×10^{-7}	40.40
4.00	8.51×10^{-7}	5.97×10^{-7}	29.89

likelihood of spontaneous detrapping at ambient temperature. This leads to more stable polarization states, particularly when combined with appropriate pulse schemes and delay intervals.

In both device types, increasing the physical size of the FeFET enhances the number of stable conductance states by incorporating more ferroelectric domains, improving analog resolution. As a result, while miniaturization may be advantageous for integration density, it must be balanced against the requirement for reliable and reproducible multistate behavior.

C. Control / Feedback

When utilizing FeFET devices as analog synapses in neuromorphic systems or in-memory computing architectures, maintaining precise conductance states is critical for ensuring computational accuracy and system reliability. However, due to process variations, pulse-to-pulse programming variability, and potential device non-idealities such as cycle-to-cycle fluctuations and polarization fatigue, the actual programmed V_{th} may deviate from the intended value. This deviation directly impacts the effective synaptic weight or memory state stored in the FeFET.

To address this challenge, a closed-loop feedback control mechanism is essential during programming. In such a scheme, each programming iteration (typically comprising SET or RESET pulses) is followed by an intermediate READ operation to evaluate the updated state of the device. If the measured conductance or V_{th} does not meet the target within an acceptable error margin, additional incremental pulses are applied until convergence is achieved [21], [34] .This adaptive approach ensures the limited number of distinct states achievable by applying identical pulses and enables fine-grained tuning of intermediate states when using incremental pulsing.

In practical hardware implementations, external measurement-based feedback (e.g., via semiconductor parameter analyzers) is not scalable. Thus, on-chip closed-loop programming circuits are integrated into FeFET arrays to enable autonomous state tracking and pulse adaptation. Common approaches include:

- **Comparator-Based Feedback:** Each FeFET synapse is connected to a local comparator that monitors its current or voltage response after each programming pulse. A digital control loop increments the pulse count or adjusts the amplitude until the comparator output indicates threshold convergence [35]–[37].
- **Successive Approximation Register (SAR):** A digital controller performs binary search over pulse amplitude or width to rapidly converge to the desired state. The synaptic state is monitored through a READ path and compared with a reference using a compact ADC or window comparator.
- **Charge Integration and Sensing Circuits:** Some architectures use charge accumulation on capacitive nodes during pulse application. The integrated charge, which correlates with ferroelectric polarization, is then sensed to infer the device state.

D. Design Insights for FeFET-Based Analog Synapses

The electrical characteristics of fabricated FeFET devices—such as achievable threshold voltage range, analog resolution under pulsed programming, retention behavior, and minimum device dimensions—significantly influence the architectural and circuit-level design of compute-in-memory (CIM) systems. This section summarizes key design insights derived from measured device behavior and programming protocols presented in this work.

a: Device Area and Integration Density

The minimum measured gate length and width of the fabricated FeFETs dictate the achievable synaptic array pitch. In this work, FeFETs with gate lengths down to *34 nm* and widths of *80 nm* were characterized. These dimensions are compatible with high-density integration in planar CMOS back-end-of-line (BEOL)-compatible processes, allowing compact array design. However, increased variability in ultra-scaled devices requires careful trade-off between density and analog weight precision. Wider gate widths are favored in arrays targeting high-resolution weight storage due to the improved signal-to-noise ratio in current-mode readout and the ability to achieve deterministic intermediate states.

VI. Conclusions

o Multi-bit operation is possible, long-term retention of intermediate states demonstrated.

o Pulse optimization to limit impacts of bias temperature instability (BTI) needs to be done.

o Deterministic or open loop operation of multi-bit FeFETs requires large-area devices with many ferroelectric grains (micron-sized lateral dimensions).

o *Something about insight into chip design choices*

REFERENCES

[1] M. Jerry, P.-Y. Chen, J. Zhang, P. Sharma, K. Ni, S. Yu, and S. Datta, "Ferroelectric fet analog synapse for acceleration of deep neural network training," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 6.2.1–6.2.4.

[2] H. Mulaosmanovic, E. T. Breyer, S. Dünkel, S. Beyer, T. Mikolajick, and S. Slesazeck, "Ferroelectric field-effect transistors based on hfo2: a review," *Nanotechnology*, vol. 32, no. 50, p. 502002, 2021.

[3] T. Mikolajick, U. Schroeder, and S. Slesazeck, "The past, the present, and the future of ferroelectric memories," *IEEE Transactions on Electron Devices*, vol. 67, no. 4, pp. 1434–1443, 2020.

[4] N. Zagni, F. M. Puglisi, P. Pavan, and M. A. Alam, "Reliability of hfo 2-based ferroelectric fets: A critical review of current and future challenges," *Proceedings of the IEEE*, vol. 111, no. 2, pp. 158–184, 2023.

[5] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature electronics*, vol. 1, no. 8, pp. 442–450, 2018.

[6] M. M. Dahan, H. Mulaosmanovic, O. Levit, S. Dunkel, S. Beyer, and E. Yalon, "Sub-nanosecond switching of si: Hfo2 ferroelectric field-effect transistor," *Nano Letters*, vol. 23, no. 4, pp. 1395–1400, 2023.

[7] X. Wang, Z. Zheng, L. Jiao, X. Chen, Y. Feng, C. Sun, Z. Zhou, D. Zhang, G. Liu, B.-Y. Nguyen *et al.*, "Unveiling the intricate dynamic charactristics of fefets with a mfmis structure: Experiment and modeling," in *2024 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2024, pp. 1–4.

[8] D. Kleimaier, S. Dünkel, H. Mulaosmanovic, J. Müller, S. Beyer, V. Havel, and T. Mikolajick, "Charge trapping and endurance degradation in ferroelectric field-effect transistors," in *2024 22nd Non-Volatile Memory Technology Symposium (NVMTS)*. IEEE, 2024, pp. 1–5.

[9] S. Dutta, C. Schafer, J. Gomez, K. Ni, S. Joshi, and S. Datta, "Supervised learning in all fefet-based spiking neural network: Opportunities and challenges," *Frontiers in neuroscience*, vol. 14, p. 634, 2020.

[10] A. T. Bartic, D. J. Wouters, H. E. Maes, J. T. Rickes, and R. M. Waser, "Preisach model for the simulation of ferroelectric capacitors," *Journal of Applied Physics*, vol. 89, no. 6, pp. 3420–3425, 2001.

[11] H. Mulaosmanovic, S. Slesazeck, J. Ocker, M. Pesic, S. Muller, S. Flachowsky, J. Müller, P. Polakowski, J. Paul, S. Jansen *et al.*, "Evidence of single domain switching in hafnium oxide based fefets: Enabler for multi-level fefet memory cells," in *2015 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2015, pp. 26–8.

[12] C.-Y. Liao, K.-Y. Hsiang, Z.-F. Lou, C.-Y. Lin, Y.-J. Tseng, H.-C. Tseng, Z.-X. Li, W.-C. Ray, F.-S. Chang, C.-C. Wang *et al.*, "Multi-peak coercive electric-field-based multilevel cell nonvolatile memory with antiferroelectric-ferroelectric field-effect transistors (fets)," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 6, pp. 2214–2221, 2022.

[13] S. Yu, J. Hur, Y.-C. Luo, W. Shim, G. Choe, and P. Wang, "Ferroelectric hfo2-based synaptic devices: recent trends and prospects," *Semiconductor Science and Technology*, vol. 36, no. 10, p. 104001, 2021.

[14] B. Zeng, C. Liu, S. Dai, P. Zhou, K. Bao, S. Zheng, Q. Peng, J. Xiang, J. Gao, J. Zhao *et al.*, "Electric field gradient-controlled domain switching for size effect-resistant multilevel operations in hfo2-based ferroelectric field-effect transistor," *Advanced Functional Materials*, vol. 31, no. 17, p. 2011077, 2021.

[15] P. Meihar, R. Srinu, V. Saraswat, S. Lashkare, H. Mulaosmanovic, A. K. Singh, S. Dünkel, S. Beyer, and U. Ganguly, "Fefet-based mirrorbit cell for high-density nvm storage," *IEEE Transactions on Electron Devices*, vol. 71, no. 4, pp. 2380–2385, 2024.

[16] T. Soliman, F. Müller, T. Kirchner, T. Hoffmann, H. Ganem, E. Karimov, T. Ali, M. Lederer, C. Sudarshan, T. Kämpfe *et al.*, "Ultra-low power flexible precision fefet based analog in-memory computing," in *2020 IEEE International Electron Devices Meeting (IEDM)*. IEEE, 2020, pp. 29–2.

[17] S. K. Gupta, D. Wang, S. George, A. Aziz, X. Li, S. Datta, and V. Narayanan, "Harnessing ferroelectrics for non-volatile memories and logic," in *2017 18th International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2017, pp. 29–34.

[18] F. Mo, Y. Tagawa, C. Jin, M. Ahn, T. Saraya, T. Hiramoto, and M. Kobayashi, "Experimental demonstration of ferroelectric hfo 2 fet with ultrathin-body igzo for high-density and low-power memory application," in *2019 Symposium on VLSI Technology*. IEEE, 2019, pp. T42–T43.

[19] S.-J. Yoon, D.-H. Min, S.-E. Moon, K. S. Park, J. I. Won, and S.-M. Yoon, "Improvement in long-term and high-temperature retention stability of ferroelectric field-effect memory transistors with metal–ferroelectric–metal–insulator–semiconductor gate-stacks using

al-doped hfo 2 thin films," *IEEE Transactions on Electron Devices*, vol. 67, no. 2, pp. 499–504, 2020.
[20] K. Ni, S. Dutta, and S. Datta, "Ferroelectrics: From memory to computing," in *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE, 2020, pp. 401–406.
[21] M. Jerry, S. Dutta, A. Kazemi, K. Ni, J. Zhang, P.-Y. Chen, P. Sharma, S. Yu, X. S. Hu, M. Niemier *et al.*, "A ferroelectric field effect transistor based synaptic weight cell," *Journal of Physics D: Applied Physics*, vol. 51, no. 43, p. 434001, 2018.
[22] S. Beyer, D. Kleimaier, S. Dünkel, H. Mulaosmanovic, S. Soss, J. Müller, Z. Jiang, K. Ni, T. Mikolajick, and H. Zhou, "Charge trapping challenges of cmos embedded complementary fefets," in *2024 IEEE International Memory Workshop (IMW)*. IEEE, 2024, pp. 1–4.
[23] S. De, F. Mueller, S. Thunder, S. Abdulazhanov, N. Laleni, M. Lederer, T. Ali, Y. Raffel, S. Duenkel, S. Mojumder *et al.*, "28 nm hkmg-based current limited fefet crossbar-array for inference application," *IEEE Transactions on Electron Devices*, vol. 69, no. 12, pp. 7194–7198, 2022.
[24] E. Yurchuk, J. Müller, S. Müller, J. Paul, M. Pešić, R. Van Bentum, U. Schroeder, and T. Mikolajick, "Charge-trapping phenomena in hfo 2-based fefet-type nonvolatile memories," *IEEE Transactions on Electron Devices*, vol. 63, no. 9, pp. 3501–3507, 2016.
[25] S. Deng, Z. Zhao, Y. S. Kim, S. Duenkel, D. MacMahon, R. Tiwari, N. Choudhury, S. Beyer, X. Gong, S. Kurinec *et al.*, "Unraveling the dynamics of charge trapping and de-trapping in ferroelectric fets," *IEEE Transactions on Electron Devices*, vol. 69, no. 3, pp. 1503–1511, 2022.
[26] H. Mulaosmanovic, S. Dünkel, J. Müller, M. Trentzsch, S. Beyer, E. T. Breyer, T. Mikolajick, and S. Slesazeck, "Impact of read operation on the performance of hfo 2-based ferroelectric fets," *IEEE Electron Device Letters*, vol. 41, no. 9, pp. 1420–1423, 2020.
[27] S. Slesazeck, H. Mulaosmanovic, M. Hoffmann, U. Schroeder, T. Mikolajick, and B. Max, "Mox in ferroelectric memories," in *Metal Oxides for Non-volatile Memory*. Elsevier, 2022, pp. 245–279.
[28] C. Alessandri, P. Pandey, A. Abusleme, and A. Seabaugh, "Monte carlo simulation of switching dynamics in polycrystalline ferroelectric capacitors," *IEEE Transactions on Electron Devices*, vol. 66, no. 8, pp. 3527–3534, 2019.
[29] D. Martin, J. Müller, T. Schenk, T. M. Arruda, A. Kumar, E. Strelcov, E. Yurchuk, S. Müller, D. Pohl, U. Schröder *et al.*, "Ferroelectricity in si-doped hfo2 revealed: a binary lead-free ferroelectric." *Advanced Materials (Deerfield Beach, Fla.)*, vol. 26, no. 48, pp. 8198–8202, 2014.
[30] T. Mikolajick, S. Slesazeck, M. H. Park, and U. Schroeder, "Ferroelectric hafnium oxide for ferroelectric random-access memories and ferroelectric field-effect transistors," *Mrs Bulletin*, vol. 43, no. 5, pp. 340–346, 2018.
[31] S. Lancaster, P. D. Lomenzo, M. Engl, B. Xu, T. Mikolajick, U. Schroeder, and S. Slesazeck, "Investigating charge trapping in ferroelectric thin films through transient measurements," *Frontiers in Nanotechnology*, vol. 4, p. 939822, 2022.
[32] X. Yin, Y. Qian, A. Vardar, M. Günther, F. Müller, N. Laleni, Z. Zhao, Z. Jiang, Z. Shi, Y. Shi *et al.*, "Ferroelectric compute-in-memory annealer for combinatorial optimization problems," *Nature Communications*, vol. 15, no. 1, p. 2419, 2024.
[33] G. Yeap, S. S. Lin, Y. M. Chen, H. L. Shang, P. W. Wang, H. C. Lin, Y. C. Peng, J. Y. Sheu, M. Wang, X. Chen, B. R. Yang, C. P. Lin, F. C. Yang, Y. K. Leung, D. W. Lin, C. P. Chen, K. F. Yu, D. H. Chen, C. Y. Chang, H. K. Chen, P. Hung, C. S. Hou, Y. K. Cheng, J. Chang, L. Yuan, C. K. Lin, C. C. Chen, Y. C. Yeo, M. H. Tsai, H. T. Lin, C. O. Chui, K. B. Huang, W. Chang, H. J. Lin, K. W. Chen, R. Chen, S. H. Sun, Q. Fu, H. T. Yang, H. T. Chiang, C. C. Yeh, T. L. Lee, C. H. Wang, S. L. Shue, C. W. Wu, R. Lu, W. R. Lin, J. Wu, F. Lai, Y. H. Wu, B. Z. Tien, Y. C. Huang, L. C. Lu, J. He, Y. Ku, J. Lin, M. Cao, T. S. Chang, and S. M. Jang, "5nm cmos production technology platform featuring full-fledged euv, and high mobility channel finfets with densest 0.021μm2 sram cells for mobile soc and high performance computing applications," in *2019 IEEE International Electron Devices Meeting (IEDM)*, 2019, pp. 36.7.1–36.7.4.
[34] C. Wang, A. Agrawal, E. Yu, and K. Roy, "Multi-level neuromorphic devices built on emerging ferroic materials: A review," *Frontiers in Neuroscience*, vol. 15, p. 661667, 2021.
[35] S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. Di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. Farinha *et al.*, "Equivalent-accuracy accelerated neural-network training using analogue memory," *Nature*, vol. 558, no. 7708, pp. 60–67, 2018.
[36] Y. Luo, P. Wang, and S. Yu, "Accelerating on-chip training with ferroelectric-based hybrid precision synapse," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 18, no. 2, pp. 1–20, 2022.
[37] S. Thomann, H. L. Nguyen, P. R. Genssler, and H. Amrouch, "All-in-memory brain-inspired computing using fefet synapses," *Frontiers in Electronics*, vol. 3, p. 833260, 2022.

VII. Authors

Sayma N. Chowdhury, photograph and biography not available at the time of publication.

Alex L. Mazzoni, photograph and biography not available at the time of publication.

Xiaohang Zhang, photograph and biography not available at the time of publication.

Andreu L. Glasmann, photograph and biography not available at the time of publication.

Sina Najmaei, photograph and biography not available at the time of publication.

Sahil Shah, photograph and biography not available at the time of publication.