PERSPECTIVE
https://doi.org/10.1038/s41928-020-00492-7

nature electronics

The future of ferroelectric field-effect transistor technology

Asif Islam Khan[1,2] ✉, Ali Keshavarzi[3] ✉ and Suman Datta[4] ✉

The discovery of ferroelectricity in oxides that are compatible with modern semiconductor manufacturing processes, such as hafnium oxide, has led to a re-emergence of the ferroelectric field-effect transistor in advanced microelectronics. A ferroelectric field-effect transistor combines a ferroelectric material with a semiconductor in a transistor structure. In doing so, it merges logic and memory functionalities at the single-device level, delivering some of the most pressing hardware-level demands for emerging computing paradigms. Here, we examine the potential of the ferroelectric field-effect transistor technologies in current embedded non-volatile memory applications and future in-memory, biomimetic and alternative computing models. We highlight the material- and device-level challenges involved in high-volume manufacturing in advanced technology nodes (≤10 nm), which are reminiscent of those encountered in the early days of high-*K*-metal-gate transistor development. We argue that the ferroelectric field-effect transistors can be a key hardware component in the future of computing, providing a new approach to electronics that we term ferroelectronics.

The ferroelectric field-effect transistor (FEFET) is a well known semiconductor device concept that until recently remained an unviable technology[1,2]. The concept appeared in a number of patents in the 1950s, and was experimentally demonstrated in the 1970s[3]. A FEFET contains a ferroelectric layer in the gate dielectric stack of a standard metal-oxide-semiconductor field-effect transistor (MOSFET). Traditionally, a FEFET is viewed as a non-volatile memory element in which binary data is stored in the direction of ferroelectric polarization (up or down). The up and down polarization directions either assist in the formation of the inversion layer in the semiconductor channel or deplete it, resulting in opposite shifts in the threshold voltage of the FEFET.

Research activities in ferroelectric devices peaked in the period from the late 1980s to the early 2000s, largely due to progress in perovskite-based, complex oxide ferroelectrics[4]. This era culminated in the successful commercialization of ferroelectric random-access memory (FRAM), in which a ferroelectric capacitor based on lead zirconate titanate (PZT) in the back-end-of-the-line (BEOL) is connected to the drain of a front-end metal-oxide-semiconductor field-effect transistor (MOSFET). Ramtron (acquired by Cypress Semiconductor), Texas Instruments, and Fujitsu marketed FRAM products for niche, low-volume applications such as smart cards, energy meters, airplane black boxes, radio frequency tags and wearable medical devices, as well as for for code storage in microcontrollers. At the same time, FEFETs garnered traction due to their attractive, non-destructive read-out functionality and energy efficiency. However, the promise of FEFETs never materialized due to the challenges involved in the integration of perovskite oxides with the front-end, semiconductor manufacturing processes—in particular, those associated with perovskite etching, hydrogen sensitivity, thickness and cell size scaling beyond the 130 nm technology node.

The discovery of ferroelectricity in hafnium oxide (hafnia) based binary oxides (fluorite structure oxides) in 2011[5,6] jumpstarted the second wave of intensive research on ferroelectric devices and, especially, FEFETs. Hafnia has been the key enabler of high-*K*-metal-gate (HKMG) technology for state-of-the-art logic transistors since mid-2000. As such, its compatibility with modern complementary metal-oxide-semiconductor (CMOS) technology and scalability can unleash the promise of FEFETs in high volume semiconductor manufacturing for a wide range of commercial products.

The revival of FEFET research is timely. Computing is evolving in a fundamentally different way, and is now being driven by data centric applications[7–9]. The traditional pillars of the semiconductor industry—dimensional scaling and the von Neumann architecture with clear separation of memory blocks and logic cores—will be insufficient to support the new ecosystem. To process massive amounts of data with high throughput and energy efficiency, the compute hardware will have to overcome the memory–logic interconnect bottleneck by adopting new, near-memory or in-memory architectures with diffused boundaries between memory and logic, while delivering continued performance gains at a rate exceeding that historically provided by scaling.

A wide range of new applications will be enabled by adding artificial intelligence (AI) to the Internet-of-Things (IoT) edge devices, a capability referred to as edge intelligence (EI)[9]. At the forefront of the data centric computing paradigm is the vision that a trillion, connected, smart edge devices will be pervasively and seamlessly integrated into the fabric of life, measuring physical parameters, processing them not at the cloud but at the edge, and making decisions in real time, leading to unprecedented opportunities for contextually intelligent applications with far-reaching societal implications. Edge intelligence will be critical for this vision because it will provide intensive, local computing at the point of data collection, thereby preventing the overburdening of communication to the central cloud by the massive collection of edge devices. AI-enabled autonomy in decision making in these small systems will also require autonomy in energy usage—the extreme need for energy efficiency—because many of these edge devices will be batteryless and will be powered by intermittent and scarce energy sources.

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [2]School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [3]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [4]Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA. ✉e-mail: akhan40@gatech.edu; akesh@stanford.edu; sdatta@nd.edu

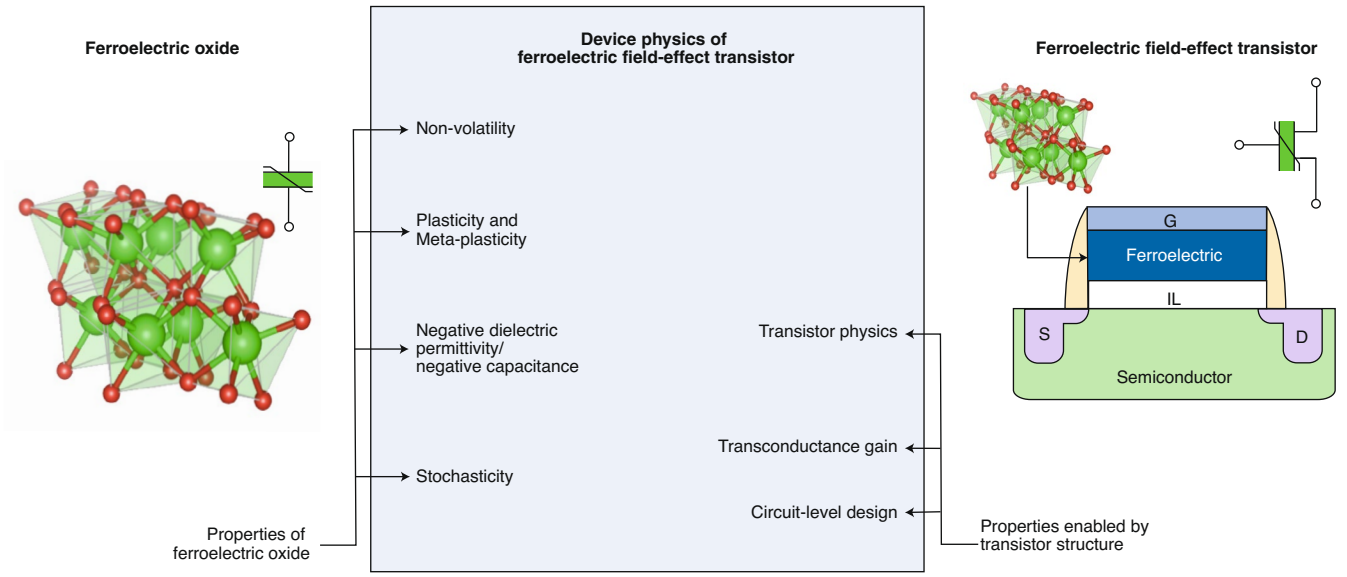


Fig. 1 | Device physics of ferroelectric field-effect transistors. A FEFET combines the rich physics of ferroelectric materials with the device physics of transistors. IL, interfacial layer.

If the semiconductor technology can support the upcoming computing paradigm effectively, the sheer number of semiconductor hardware chips required could drive the next stage of exponential growth of the semiconductor industry, transforming it toward an over-a-trillion-dollar business. This motivates a bottom-up approach, innovating at the lowest level of the computing hierarchy—the material and device level—to deliver the required functionalities beyond what is available with current CMOS platforms. Such innovations can then be leveraged at the circuit, micro-architecture, system and software level to deliver autonomy by energy efficient computing, logic–memory colocation, and performance gain while improving performance per watt.

In this Perspective, we examine the FEFET technology in the context of emerging data-centric applications. We establish the key device-level challenges for the commercial viability of this technology. We argue that FEFETs, with their energy and area efficiency and diverse merged logic–memory functionalities, will be the forerunner of a new approach in electronic devices, which we refer to as ferroelectronics, and that this approach will be critical in addressing the future needs of computing.

Ferroelectric device physics leads the way

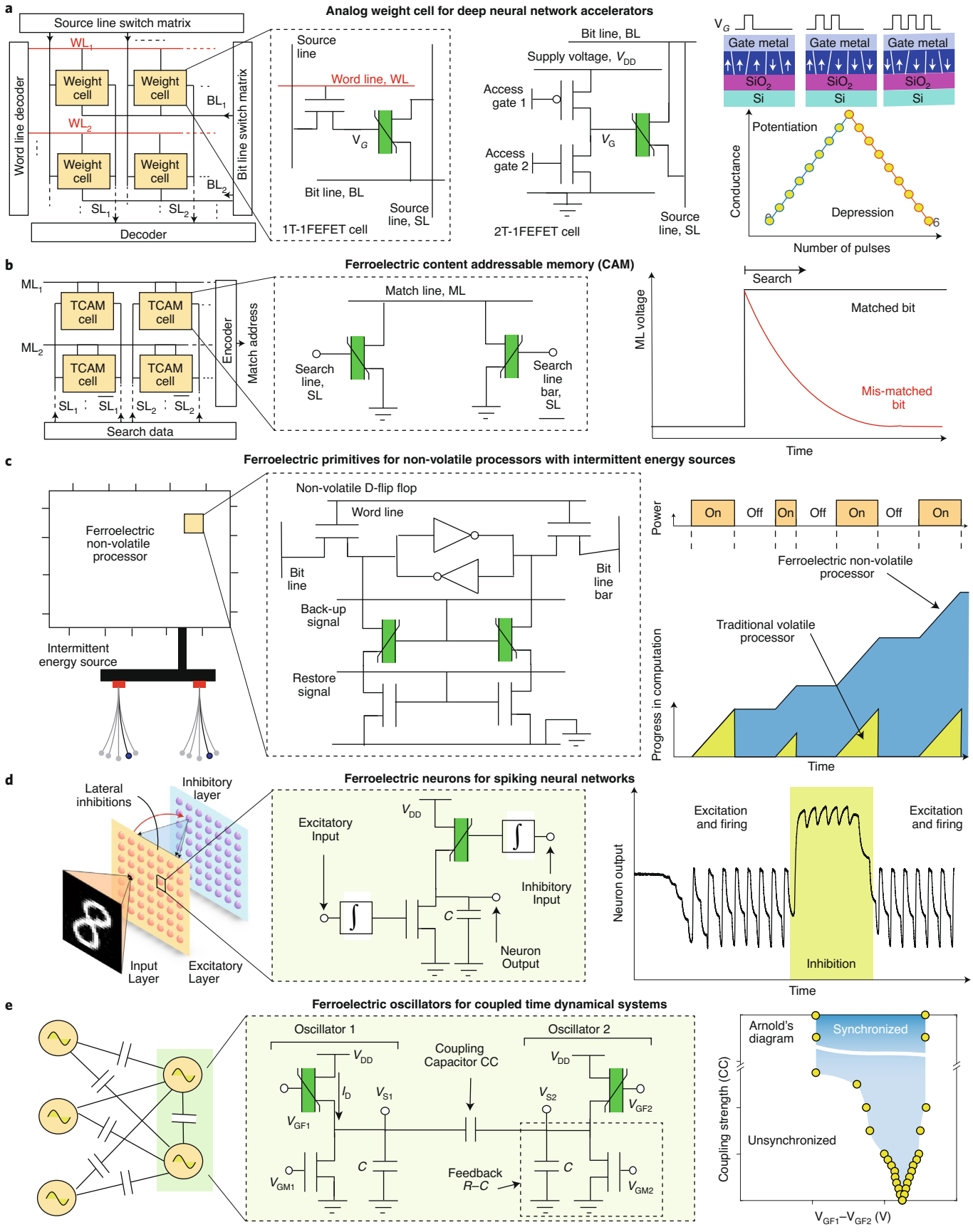
Ferroelectric materials exhibit unique features such as hysteresis, non-volatility, plasticity, negative dielectric permittivity/negative capacitance[10], stochasticity[11], ferroelasticity and multiferroicity[12], semiconduction[13] and even nonlinear and chaotic behaviour[14] and quantum mechanical effects[15]. In a FEFET, the intrinsic ferroelectric dynamics is strongly coupled to the conductance state of the underlying semiconductor channel (Fig. 1). Being a three-terminal device with a transconductance, FEFETs allow for a wide gamut of circuit designs that leverage the unique ferroelectric physics to efficiently address the diverse needs of traditional and emerging computing applications.

As such, while a FEFET is primarily viewed as a non-volatile memory element, its portfolio has expanded significantly beyond memory applications and now includes negative capacitance transistors for ultra-low power, high performance logic technology[10], analogue weight cells for deep neural networks (DNNs) and in-memory computing[16–20], content-addressable memory cells for fast and highly parallel database search and finding match locations[21], artificial neurons for spiking neural networks (SNNs)[22,23], coupled oscillatory networks for continuous time dynamical systems[24,25], circuit primitives for stochastic computing, fast data back-up and wake-up circuits for intermittent computing[21] and so on—all of which utilize combinations of these ferroelectric features and characteristics[26]. In this section, we analyze how the device physics of the FEFET relate to different data-centric applications (Fig. 2), except for negative capacitance, a summary of which can be found in ref. [10].

Ferroelectric plasticity—that is, the stable, partially switched states in the ferroelectric layer programmed by sub-coercive

Fig. 2 | Applications of ferroelectric field-effect transistors and the corresponding device physics. a, Multi-state weight cell/analogue synapse for deep neural network accelerators. Plasticity of the ferroelectric polarization with sub-coercive voltages leads to multiple conductance states in FEFETs. Linear and symmetric potentiation and depression characteristics with identical pulses can be obtained in a two transistor, one FEFET (2T1FEFET) cell[20]. **b**, Ferroelectric ternary content addressable memory (TCAM)[28,29]. The match-line (ML) stays charged when the bit placed on search line matches the stored bit, while ML voltage discharges in case of mismatched bit. **c**, Ferroelectric non-volatile processor (FE-NVP) powered by an intermittent and unreliable energy source. The D flip flop (DFF) states are backed up in the FEFETs when power failures are anticipated, which is controlled by the back-up signal. When the power is restored, the data from the FEFETs is restored into the DFF controlled by the restore signal. With frequent power outages, computation progresses seamlessly in an FE-NVP while in a traditional volatile processor, computation has to be rolled back every time a power outage occurs. **d**, Ferroelectric neurons for spiking neural network applications[22]. The anti-plastic behaviour (the lack of stable, intermediate states in the hysteretic jumps in FEFET drain current) allows for relaxation oscillations in a simple two transistor circuit topology. The ferroelectric neuron can accommodate both inhibitory and excitatory functionalities. **e**, Ferroelectric oscillators for coupled time dynamical systems[24]. The synchronization dynamics is depicted in the Arnold's tongue diagram which plot the range of the gate voltage difference (V_{GF1}–V_{GF2}) for oscillator synchronization for given values of the coupling strength (C_c). The coupled dynamics of ferroelectric oscillators can be utilized to solve computationally hard problems such as graph coloring, convex optimization and so on.

voltages—leads to multi-state, non-volatile operation in FEFETs. This functionality is particularly important creating ultra-dense memories—for standard embedded memory applications and as multi-state weight cells or 'analogue' synapses in deep neural network accelerators. Multi-bit operation with 2–8 bits (4–256 levels), order of 100-fold conductance modulation, ~5 ns update pulses,



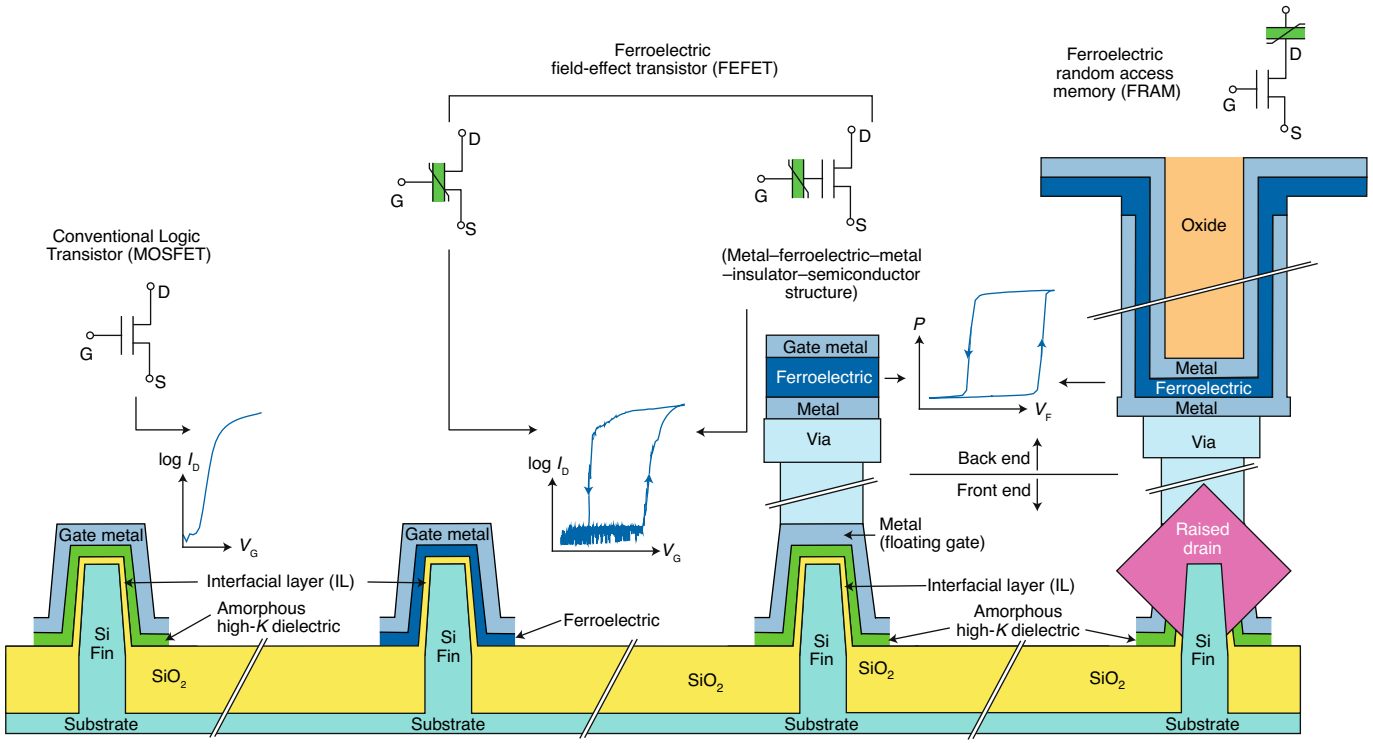


Fig. 3 | Embedded ferroelectric memory technologies. a, Schematic representation of a conventional complementary metal–oxide–semiconductor (CMOS) logic transistor and ferroelectric field-effect transistors (FEFETs) and a ferroelectric random-access memory (FRAM). Two different constructs of FEFETs with metal–ferroelectric–insulator–semiconductor (MFIS) and metal–ferroelectric–metal–insulator–semiconductor (MFMIS) structures are shown. Unless otherwise noted, a FEFET would refer to the one with a MFIS structure. I_D, V_G, P and V_F refer to the transistor drain current, the transistor gate voltage, the ferroelectric polarization and the voltage across the ferroelectric capacitor, respectively.

and linear and symmetric potentiation–depression with identical pulse sequences has recently been demonstrated in FEFETs[16–20]. With deep learning models adding more layers and requiring more parameters, device-level requirements for implementing 'analogue' weight cells have become extremely stringent especially for high-accuracy training in area and energy efficient, deep neural network (DNN) accelerators[27]. While none of the existing and emerging memory technologies meet these requirements, these early FEFET results indicate the possibility that FEFETs may be able to meet the ideal analogue weight cell characteristics.

The three-terminal device architecture of FEFET allows for unique circuit topologies in specialized memory applications. For example, a FEFET can act as both the selector and the non-volatile memory element for ternary content addressable memories (TCAMs) leading to the smallest footprint TCAM cell with just two transistors[28,29]. With superior array-level performance compared to other technologies such as magnetic tunnel junction (MTJ), resistive RAM (RRAM), and even SRAM, ferroelectric TCAMs can enable traditional high-speed data-processing applications such as internet protocol filters and network routers as well as beyond deep neural network applications such as one/few shot learning in memory augmented neural networks (MANN)[28].

Another application of fine-grained integration of FEFETs with standard CMOS logic is in intermittent computing. Especially in edge intelligence applications, where the energy source is intermittent and unreliable (such as solar, ambient radio frequency, vibration, and thermal energy), the intermediate computation states in the pipelining logic need to be backed up into on-chip non-volatile memory frequently in anticipation of unpredictable power failures, and restored when the power returns[9]. FEFETs with their superior energy profile and high speed can enable the most energy and area efficient circuit primitives for intermittent computing (such as non-volatile latches, registers and SRAMs)[30]; for example, non-volatile D flip flops (DFFs) design has recently been demonstrated by adding just two FEFETs to its conventional CMOS counterpart[21]. Such FEFET-enabled systems can also be utilized for scheduled power gating in both low-power mobile applications and high-performance server processors to reduce static leakage power.

At scaled dimensions, FEFETs exhibit unusual effects: meta-plasticity, single-domain-like switching and stochasticity. Being laterally confined to sub-100 nm with only a few structural grains present, the switching dynamics in scaled FEFETs are dominated by nucleation rather than domain growth and coalescence—unlike that in large-area ferroelectric capacitors. In case of meta-plasticity, which is also termed polarization accumulation, ferroelectric polarization switches progressively through local nucleation, and acts as a hidden, continuous variable resulting only in a single-domain-like, abrupt jumps in the conductance[31]. In a narrow region of phase space of program voltage and time, conductance switching in scaled FEFETs is not only activated by a critical threshold crossing but also become probabilistic[31]. In addition to enabling multi-bit, embedded memory with scaled dimensions, these phenomena together can allow for ultra-dense, energy-efficient, and massively parallel correlation detection systems for real-time sensor analytics[32], on-chip random number generation for security applications[33] and leaky-integrate-and-fire neurons[31,34].

The opposite of plasticity effect can also be observed in FEFETs—that is, antiplasticity: the lack of stable, intermediate state in the hysteretic jumps of the FEFET drain current. This occurs when a metal layer is introduced to sandwich the ferroelectric layer to form a metal–ferroelectric–metal–insulator–semiconductor (MFMIS) structure in a FEFET. Note that a typical FEFET uses a metal–ferroelectric–insulator–semiconductor (MFIS) structure where the ferroelectric layer is not in between metal layers. Antiplasticity can

Table 1 | Key parameters and metrics

Metrics	Mainstream embedded memory					Embedded ferroelectric memory			
	eSRAM	eDRAM[70]	eFlash (FG)	eFlash (SG MONOS)[42]	eFlash (SONOS)[41]	FEFET (hafnia based, MFIS structure)	FEFET (hafnia based, MFMIS structure)	FRAM (hafnia based)	FRAM (perovskite based)[67]
Cell size	120–150F^2	40F^2	40–60F^2	40–50F^2	50–60F^2	10–30F^2	10–30F^2	30–40F^2	50–60F^2
Cell structure	6T	1T1C	1.5T	1.5T	2T	1T	1T1FE, 1T	1T1FE, 2T2FE	1T1FE, 2T2FE
Non-volatile	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Multi-bit operation	No	No	Yes	Yes	Yes	Yes	Yes	No	No
Non-destructive read	Yes	No	Yes	Yes	Yes	Yes	Yes	No	No
Status	Av.	Dev.	Av.	Dev.	Dev.	Res.	Res.	Res.	Av.
Advanced node demonstration	7 nm FinFET	22 nm FinFET	40 nm	16 nm FinFET	28 nm HKMG	22 nm FDSOI	N/A	N/A	130 nm
Write voltage	<1 V	<1 V	~12 V	~12 V	~5 V	1.5–4 V	~1.5 V	1–3 V	1.5 V
Write energy	~1 fJ	~1 pJ	~100 pJ	~100 pJ	1–10 pJ	1–10 fJ	1–10 fJ	~100 fJ	~1 pJ
Standby power	High	Medium	Low	Low	Low	Low	Low	Low	Low
Write speed	<1 ns	>10 ns	~100 ns	<100 ns	~100 ns	1–10 ns	1–10 ns	1–10 ns	1 µs
Read speed	<1 ns	>10 ns	~10 ns	<10 ns	~10 ns	1–10 ns	1–10 ns	1–25 ns	50–100 ns
Endurance	>10^{16}	>10^{16}	~10^4	~10^5	~10^6	10^5–10^9	>10^{10}	>10^{12}	>10^{14}

Key device parameters and performance metrics comparing current embedded memory candidates and ferroelectric technologies. Data for eDRAM, SG MONOS eFlash, SONOS eFlash and perovskite based FRAM are obtained from ref. [70], ref. [42], ref. [41] and ref. [67], respectively. FG, floating gate; SG MONOS, split gate metal–oxide–nitride–oxide–Si; SONOS, Si–oxide–nitride–oxide–Si; eSRAM, embedded static random-access memory; eFlash, embedded flash; eDRAM, embedded dynamic random-access memory; FRAM, ferroelectric random access memory; T, transistor; C, capacitor; FE, ferroelectric; Av., commercially available; Dev., development; Res., research.

enable biomimetic computing models such as spiking neural networks[22], coupled oscillatory time dynamical systems[24] and swarm intelligent networks[25]. At the heart of these compute models is a relaxation oscillator that encodes information in time domain—in the frequency, phase, spike timing or spike count. One of the unique features of the ferroelectric neural primitives is that they can accommodate inhibitory functionality in addition to its excitatory counterpart in a compact, two transistor topology—which is one of the key requirements for successful learning and high-accuracy inferencing in spiking neural networks[22]. This is in contrast with CMOS neurons that require tens of transistors, and alternative emerging materials/devices-based neurons which do not have the built-in inhibitory function. In addition, ferroelectric spiking neurons have characteristic features beyond simple leaky-integrate-and-fire dynamics and can exhibit a large number of dynamical phenomena of cortical neurons unlike CMOS and alternative memory-based implementations[23]. In addition, spiking or oscillatory neurons with inhibitory and excitatory inputs—with recurrent connectivity based on analogue, ferroelectric synapses—can allow for efficient implementation of reservoir computing models for temporal data classification and pattern generation.

Vertical FEFETs in dense, three-dimensional (3D) NAND architecture are being actively pursued[35]. This is interesting from a stand-alone-memory point of view due to their potential for low voltage operation compared to that of the mainstream 3D NAND flash technology. In addition, interesting concepts of FEFET based reconfigurable memory devices have recently been proposed in which the hysteresis can be dynamically tuned and the operating mode can be switched from 'non-volatile' to 'volatile' on the fly by an additional control terminal[36,37]. Such reconfigurablity can, in principle, broaden the design space for non-volatile memory and neural network applications. Ferroelectricity in hafnia-based binary oxides is robust at both high and cryogenic temperatures and even in radiation environments[38,39], which makes FEFETs attractive as an extreme-environment memory technology.

The memory landscape

We now address the question: where does FEFET fit in the embedded memory space? The game-changer for all data centric applications is high-bandwidth, low-latency, energy-efficient, dense embedded memory that is programmable with low, logic-compatible voltages (<1.5 V). Unlike stand-alone memories, embedded memory shares the same silicon as its logic counterpart and, hence, enable efficient data transfer between logic cores and memory units in traditional von Neumann computers and can support upcoming, in-memory computing architectures. Today, embedded static random-access memory (eSRAM) is the only commercially viable embedded memory solution available on advanced technology nodes (sub-28 nm), which, however, is volatile and suffers from a rather large cell size (>120–150F^2, F being the feature size) and high standby static leakage power. Embedded DRAM (eDRAM) is also volatile, and faces the challenges of high power dissipation due to the need to frequent refresh operations. In the embedded non-volatile memory (eNVM) space, the workhorse is the embedded flash (eFlash) technology based on floating gate (FG) transistors[40]. However, the scalability of the FG eFlash technology into 28 nm and sub-28 nm nodes will arguably hit a wall due to both technical and economic challenges, the most important ones being increased mask counts (>13) and hence, the cost, the requirement of very high voltages (~12 V), limited endurance (~10^4 cycles) and the unavailability of polysilicon in the HKMG technology to implement the floating gate[9]. The potential 'mainstream' alternative to floating gate technology is the oxide–nitride–oxide (ONO) based eFlash where the polysilicon FG is replaced with a nitride charge trapping layer. ONO-eFlash technologies have higher endurance (10^5–10^6 cycles), and is integratable with advanced nodes with demonstrations at the 28 nm HKMG node in the Si–oxide–nitride–oxide–Si (SONOS) structure[41] and the 16 nm FinFET node in the split-gate metal–oxide–nitride–oxide–Si (SG MONOS) structure[42], respectively. SONOS eFlash devices can operate at smaller voltages (5–7 V) due to gentler Fowler–Nordheim tunneling mechanism, and require a lower number of additional

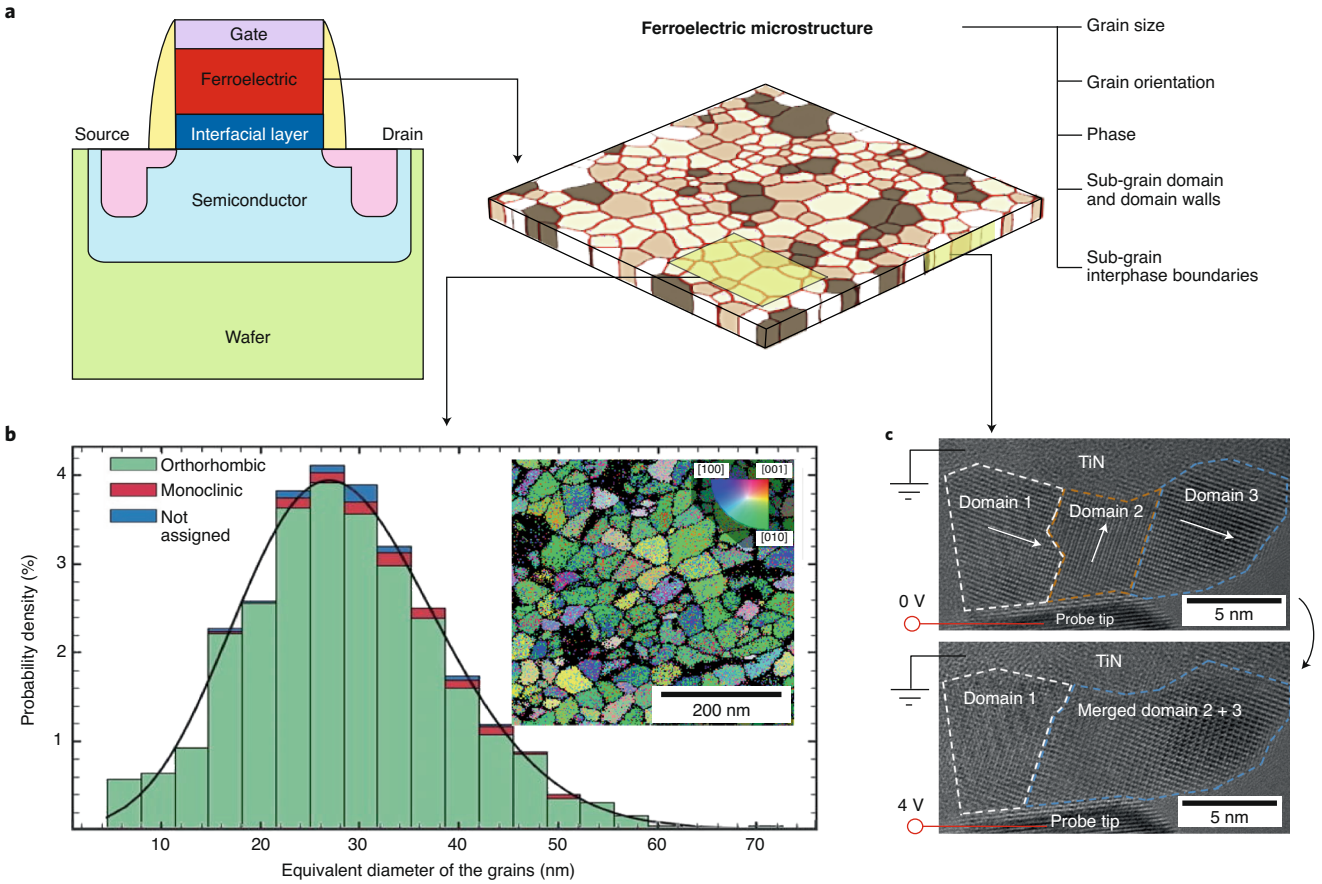


Fig. 4 | Origin of threshold voltage (V_{th}) variation in scaled ferroelectric field-effect transistors. a, Due to ferroelectric microstructure in an FEFET is disordered and nano-crystalline and varies spatially in terms of grain size, phase distribution (ferroelectric orthorhombic $Pca2_1$, non-ferroelectric monoclinic $P2_1/c$ and tetragonal $P4_2/nmc$ phases), orientation, grain boundaries, sub-grain domain walls and interphase boundaries and so on. Given that the modal radius of the grains is of the order of 10 nm, a scaled FEFET may contain a few randomly distributed structural grain leading to device-to-device V_{th} variability. **b**, Grain size distribution of the monoclinic and orthorhombic phases in a 10 nm $Hf_{0.5}Zr_{0.5}O_2$ thin film[52]. The equivalent diameter of the orthorhombic grains is between 19 and 70 nm with a mean value of ~29 nm, while the monoclinic grains have a mean grain size of about 34 nm. The inset shows grain orientation along the *z* axis of the sample extracted from the Kikuchi pattern by color coding of the unit cell axis which is parallel to the respective sample-axis. Many grains are present with the [010]-axis pointing out-of-plane (along the *z* axis), thus indicating a nominal film texture. **c**, The evolution of microstructure in an antiferroelectric ZrO_2 film with an applied voltage measured using in-situ high resolution transmission electron microscopy[71]. With a 4 V bias applied to the probe-tip, the two right-most domains coalesce to form a larger domain in ZrO_2. Reversible and irreversible microstructural changes during program/erase cycles along with charge trapping mechanisms can lead to cycle-to-cycle variation in scaled FEFETs. Figure reproduced with permission from: **b**, ref. [52], AIP; **c**, ref. [71], IEEE.

masks (~5). However, the voltage is still not logic compatible, and hence, they require charge-pump circuits leading to degraded array efficiency and lower memory density.

To fill this gap, several emerging candidates are being extensively investigated for the better part of two decades—such as spin transfer torque magnetic random-access memory (STT-MRAM), resistive random-access memory (RRAM), and phase change memory (PCM). In fact, several industrial houses have demonstrated early prototypes of these emerging eNVMs with back-end-of-the line integration at relatively scaled technology nodes including: STT-MRAM on the 22 nm and the 28 nm platforms[43,44] and RRAM on the 22 nm node[45] and PCM on the 40 nm node[46]. However, STT-MRAM, RRAM and PCM—all being current-driven devices—are bottlenecked by their high write energies. STT-MRAM has a large cell size due to the large area of the access transistor required for providing a high write current, lacks multi-bit functionality and requires deposition of a large number of precisely controlled thin film layers. RRAM suffers from inherent variability in filament formation process.

A relatively newcomer in this crowded field of embedded memory, FEFET offers at least three distinctive advantages. First, the close resemblance between FEFET and the HKMG transistor technology (Fig. 3 and Table 1) suggests that FEFETs have the same scalability as that of the state-of-the-art logic transistors down to sub-10 nm nodes with a greatly reduced, 'additional' mask count (≤2) compared to eFlash technologies as well as other non-volatile memories. To that end, FEFETs have already been integrated in 28 nm planar bulk CMOS and 22 nm fully depleted planar silicon-on-insulator (FDSOI) CMOS platforms as an embedded memory technology with array-level demonstrations[47,48]. Second, and equally impressive, is the fact that the energy profile of FEFET is the best-in-class among all non-volatile memory technologies, approaching the realm of volatile eSRAM. Being electric-field driven, ferroelectricity provides a write energy in FEFET that is at least two or three orders of magnitude lower than that in STT-MRAM, RRAM and PCM and is within an order of magnitude of that of eSRAMs. Third, the transistor action in FEFETs, which is not available in other two terminal memories, allows not only for fast, non-destructive read but also

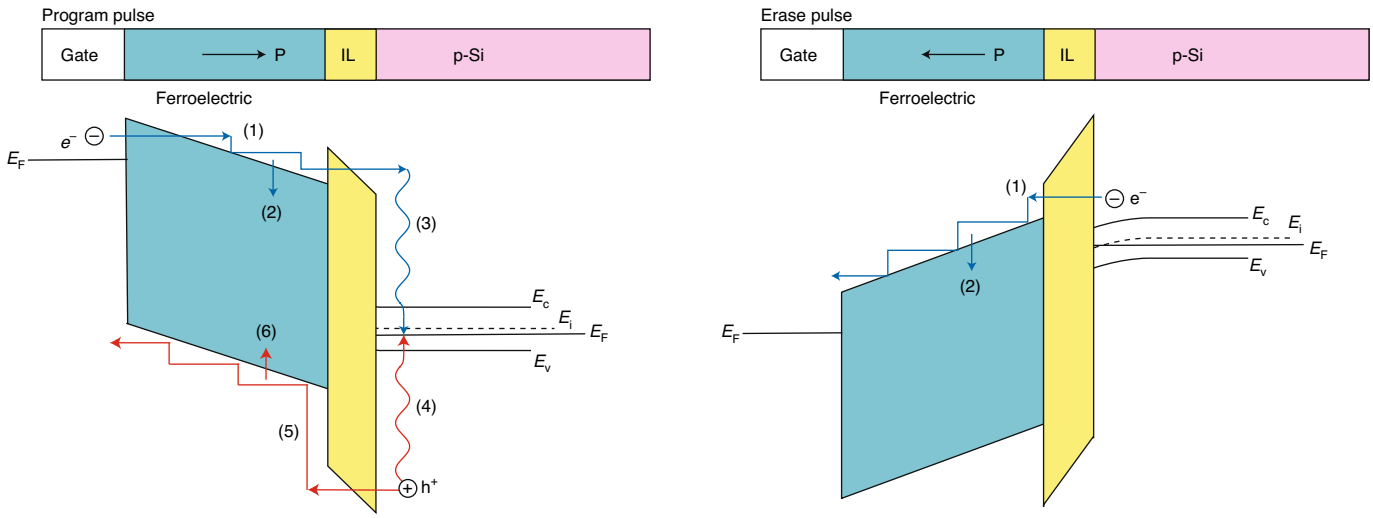


Fig. 5 | Origin of endurance limitation in ferroelectric field-effect transistors. Energy band diagrams of the gate metal–ferroelectric–IL–semiconductor cross-section in an FEFET during program and erase cycles are shown. In both cases, the band diagrams correspond to the initial epochs during which the polarization has not yet switched in response to the voltage pulse. e^- and h^+ represent electrons and holes, respectively. E_F, E_c, E_v and E_i refer to the Fermi level (of either the semiconductor or the gate metal), the conduction band edge, the valence band edge and the intrinsic Fermi level of the semiconductor, respectively. The arrows in blue and red represents electron and hole transport, respectively, through the gate stack. The curly, vertical lines represent impact ionization processes. Different processes that are potentially responsible for FEFET degradation are denoted by[60–64]: (1) electron injection from the cathode; (2) electron trapping in the bulk, the interface and the grain boundaries; (3) energy released by hot electrons and trap generation at the Si-SiO_2 interface; (4) hot hole generation by impact ionization and injection into the gate stack; (5) energy released by hot holes and trap generation at the FE-SiO_2 interface; (6) hole trapping and trap generation in the bulk and the grain boundaries. Continued program and erase operations lead to charge trapping and trap generation in and degradation of the quality of the gate dielectric stack resulting in a progressive shift of the threshold voltages and a gradual narrowing of the memory window.

unique, efficient and creative cell and circuit designs with low-area footprint (10–30F^2 depending on application). In addition, the write operation in ferroelectric devices can be extremely fast and less than 1 ns (ref. [49]).

The technological challenges

The path forward for FEFETs in the domain of data-centric applications definitely involves porting them into advanced technology nodes at 10 nm and beyond—in FinFET and even nano-sheet architectures. This is because scaling not only increases the embedded memory density but also improves its raw performance such as read/write speed and, most importantly, provides compatibility with high-performance logic located on the same chip/die. For the ferroelectric gate stack to fit into the tight pitch at such nodes, the ferroelectric layer needs to be scaled from the current state-of-the-art of 5–10 nm to 3 nm and below. To that end, robust ferroelectricity in hafnia-based oxides has already been demonstrated at these scaled thicknesses[50,51]. Ferroelectric thickness scaling can also lead to the reduction of the write voltage from the current 2–4 V to on-chip logic compatible voltages (≤1.2 V), thereby eliminating the need for charge pump circuits and hence better memory array efficiency (albeit at the cost of a reduced memory window).

For array-level performance at advanced nodes, the device-to-device variation of the threshold voltage (V_{th}) is an important consideration. Due to the lack of an epitaxial template for a ferroelectric on a semiconductor in a FEFET, the ferroelectric microstructure is mesoscopically disordered and polycrystalline, varying spatially in terms of grain radius (2–50 nm), phase distribution (ferroelectric orthorhombic $Pca2_1$, non-ferroelectric monoclinic $P2/c$ and tetragonal $P4_2/nmc$ phases), orientation, grain boundaries, sub-grain domain walls and interphase boundaries[52,53] (Fig. 4). A scaled FEFET may contain a few randomly distributed structural grains, which is an intrinsic source of V_{th} variability—in addition to traditional sources such as random dopant fluctuations, fabrication tolerances and so on[54]. This bears similarity to the metal gate polycrystallinity induced work-function variation in early HKMG technologies. By making the metal layers extremely nanocrystalline (1–2 nm of grain size) or 'pseudo-amorphous', the device-to-device V_{th} variation was alleviated in HKMG transistors. It is possible that a similar grain engineering approach could be adopted for FEFETs: if the average ferroelectric grain size can be reduced to 2–3 nm while still preserving ferroelectricity within, each of the scaled FEFETs would contain hundreds of grains. In such a case, non-uniformities in microstructure average out within the device itself leading to a reduced device-to-device variation. Conversely, engineering structural texture in rather large ferroelectric grains[55] or converting amorphous layers altogether into single crystal-like ones[56] can be powerful approaches to control V_{th} variation. Nonetheless, better insights into the FEFET variability aspects will require fundamental materials-level studies of voltage-induced microstructural evolution at both microscopic (atomic) and mesoscopic scales (grain structures), potentially employing in-situ and multi-scale structural probing techniques[57].

State-of-the-art FEFETs have a field-cycling endurance in the range of 10^5–10^9 cycles for deterministic switching[47,48,57,58]. This endurance, while being significantly better than that in eFlash and similar to those in STT-MRAMs and RRAMs, is limited compared to that of eSRAM. It is to be noted that a hafnia-based ferroelectric oxide in a metal–insulator–metal structure can endure much greater number of cycles of deterministic switching (~10^{11})[59]. In ferroelectric capacitors, effects such wake-up, ageing, imprint and fatigue are related to multiple microscopic mechanisms including oxygen vacancies, grain boundaries and structural defects, carrier injection, domain and domain-wall pinning, structural phase change and so on. In a FEFET, additional degradation mechanisms such as charge trapping and trap generation at the interfaces, and potentially electron–hole recombination and hot carrier effects are present[60–64]. In fact, similar to the degradation aspects of HKMG technology, the

FEFET endurance limitation is predominantly an interface issue: The weakest link in a FEFET is not the ferroelectric layer itself, but rather the interfaces in the metal–ferroelectric–insulator–semiconductor structure[60–64]. Due to its lower dielectric constant (~4), most of the gate voltage drop during write operation appears across the interfacial oxide layer (IL) (Fig. 5); this leads to an IL electric field as high as 10–15 MV cm^{-1} (ref. [61]). Improving the endurance will require a meticulous gate stack design that improves the interface quality and increases the voltage drop across the ferroelectric layer and decreases it across the interfacial oxide layer. The known techniques in the HKMG technology on reducing the IL thickness via scavenging techniques, increasing the dielectric constant of the IL, and passivation of bulk, interface and grain boundary trap need to be re-explored in the FEFET context[65]. Reducing the write voltage will reduce the hot carrier related degradation mechanisms in FEFETs as well. Interestingly, unlike the case in FRAMs, relatively low values of spontaneous polarization and dielectric constant of ferroelectric layer can be conducive for FEFETs[60]. An alternative strategy to improve endurance is to eliminate the ferroelectric-IL interface altogether in a FEFET with a MFMIS structure that contains an intermediate, floating metallic gate with the ferroelectric layer integrated in the BEOL. In fact, endurance greater than 10^{10} cycles has recently been demonstrated in such FEFET structures[19].

Ultimately, technological development of FEFETs will essentially be a trade-off among all metrics: endurance and reliability, memory window and on–off ratio, scalability, latency/access time and operating voltage, retention time, variability and so on. The design space of FEFETs will not have any singular region of optimality; it will rather be dictated by the specific requirements of data-centric applications and perhaps based on the level of data persistence/retention that the application demands. For example, plasiticity/partial polarization switching in FEFETs enables desirable features such as dense multi-bit-per-cell operation, low and logic-compatible write voltage and potentially unlimited endurance at the expense of reduced memory window and on–off ratio, decreased memory retention and increased V_{th} variation[61,66]. IL scaling will adversely affect the channel mobility while ferroelectric thickness scaling will reduce the memory window, both of which, in turn, increases the memory access time. The good news is that many of the current challenges of the FEFET technology are reminiscent of those in the early days of the HKMG technology from the mid-1990s and into the 2000s, which ultimately became one of the glowing success stories of the semiconductor industry.

At this point, it is worth noting that FRAM can be positioned as another embedded memory technology which, being enabled by ferroelectricity, shares some of the attractive features of FEFETs such as energy efficiency and non-volatility. Lead zirconate titanate (PZT) based FRAMs with 1T1C (one transistor, one capacitor) cell configuration are commercially available as embedded memories for microcontrollers and digital signal processors[67] and stand-alone NOR memory solutions[68]. However, due to the 'lack' of the transconductance gain that FEFET offers, FRAM requires a destructive read, is slower and read sensing requires an accurate reference which adds complexity to cell design, perhaps often requiring a 2T2C configuration. This limits density and array efficiency as well as the design space for circuit-level motifs. On the other hand, FRAM potentially enjoys superior endurance (due to the lack of IL and semiconductor interfaces) and voltage scalability compared to those in their FEFET counterparts, and hence is particularly well suited for DRAM-type applications.

FEFETs can dominate the upcoming data-centric computing paradigm by being complementary to eSRAM and CMOS based volatile memories and by being an alternative with superior density and energy efficiency in application specific, hierarchical memory architectures. This is because applications in this new paradigm range from the cloud to the client to the edge varying dramatically in terms of power, performance, area and cost; no single, embedded technology can span the entire space. For example, dedicated deep neural network inference engines in the cloud will require ultra-dense memories with ultra-fast access times. In such applications, high performance, advanced node FEFETs (potentially even in a monolithic 3D configuration on top of CMOS logic or in a 3D vertical column like SONOS based 3D NAND structure) with modest endurance can be the key enablers. Autonomous systems with lifelong and in-field learning demands will require frequent write into memory banks with relaxed data retention time, and hence, endurance traded off with retention with partial polarization switching will be appropriate. In the paradigm of edge intelligence with a trillion, connected smart devices, performance with brutal energy efficiency will be the most important requirement. Especially when powered by scavenged/intermittent power sources, the back-up and restore functionality will require intricate meshing between non-volatile technology with the CMOS-based primitives, that is, eSRAM, register, latches and flip-flops. In moving AI training to the edge nodes (referred to as small-system AI), the neural networks are likely to be partly pre-programmed requiring frequent, in-the-field weight updates in only a few dynamic layers enabled by ultra-fast, high endurance eSRAMs while most others will be static enabled with a non-volatile memories with high retention and moderate endurance—that is, FEFETs[9]. Altogether, advanced node FEFETs provide non-volatility, a best-in-class energy profile and unique cell and circuit design capabilities, while complementary eSRAMs enable ultra-fast speed and raw performance. Hence, in the future, data centric computing paradigm, the FEFET-eSRAM combination will enable architecting application specific, fluid memory hierarchies that balance energy, density, performance, memory access patterns, area and cost.

The era of ferroelectronics

Over the last century, ferroelectricity has yielded a remarkable range of fundamental discoveries and innovations[69], which, from a technology standpoint, culminated in the successful commercialization of a memory product with one of the best energy profiles: FRAMs based on perovskite oxides. With the discovery of ferroelectricity in modern CMOS-compatible and scalable materials—hafnium oxide and zirconium oxide and their alloyed variants—the stage is set for ferroelectrics to come to the forefront of mainstream semiconductor electronics, creating an era of ferroelectronics.

Ferroelectronics is not a singular technology. Instead, it uses a diverse set of devices and elements that include negative capacitance logic transistors, non-volatile memory devices such as FEFETs and FRAMs, magnetoelectric–multiferroic devices, and foundational technologies such as CAM, analogue and radio-frequency capabilities[9]; FEFETs are central to all of this. In our view, FEFETs are not just yet another memory device. They are rather a memory–compute element and one of most versatile transistor technologies ever conceived, thanks to their rich device physics. The progression of FEFETs will follow a path similar to that of the HKMG transistor, which, to date, has been one of most successful technologies in semiconductor electronics.

Data-centric computing will require a pervasive presence of ferroelectronics and most importantly, FEFETs. These devices offer performance, energy and area efficiency, and unique, merged logic–memory functionalities, and can complement CMOS-based logic and volatile memory technologies in supporting diverse workloads and applications. The next wave of exponential growth of the semiconductor industry depends on how well the electronic hardware can support the explosion of data-centric computing applications, and we believe that the FEFET technology will be a key component in delivering this future.

Received: 22 April 2020; Accepted: 24 September 2020;
Published online: 19 October 2020

References

1. Ma, T. & Han, J.-P. Why is nonvolatile ferroelectric memory field-effect transistor still elusive? *IEEE Electron Device Lett.* **23**, 386–388 (2002).
2. Mikolajick, T., Schroeder, U. & Slesazeck, S. The past, the present, and the future of ferroelectric memories. *IEEE Trans. Electron Devices* **67**, 1434–1443 (2020).
3. Sugibuchi, K., Kurogi, Y. & Endo, N. Ferroelectric field-effect memory device using $Bi_4Ti_3O_{12}$ film. *J. Appl. Phys.* **46**, 2877–2881 (1975).
This work demonstrated one of the earliest ferroelectric field-effect transistors and used a perovskite oxide based ferroelectric.
4. Scott, J. F. *Ferroelectric Memories* Vol. 3 (Springer, 2000).
5. Böscke, T. et al. Phase transitions in ferroelectric silicon doped hafnium oxide. *Appl. Phys. Lett.* **99**, 112904 (2011).
6. Böscke, T., Müller, J., Bräuhaus, D., Schröder, U. & Böttger, U. Ferroelectricity in hafnium oxide: CMOS compatible ferroelectric field effect transistors. In *2011 Int. Electron Devices Meeting* 24.5.1–24.5.4 (IEEE, 2011).
This work demonstrated a ferroelectric field-effect transistor with a hafnium oxide based ferroelectric.
7. Salahuddin, S., Ni, K. & Datta, S. The era of hyper-scaling in electronics. *Nat. Electron.* **1**, 442–450 (2018).
8. Aly, M. M. S. et al. Energy-efficient abundant-data computing: The N3XT 1,000x. *Computer* **48**, 24–33 (2015).
9. Keshavarzi, A. & van den Hoek, W. Edge intelligence—on the challenging road to a trillion smart connected iot devices. *IEEE Des. Test* **36**, 41–64 (2019).
10. Wong, J. C. & Salahuddin, S. Negative capacitance transistors. *Proc. IEEE* **107**, 49–62 (2018).
11. Diestelhorst, M. & Drozhdin, K. Stochastic resonance and domain switching. *Ferroelectrics* **291**, 217–224 (2003).
12. Heron, J. et al. Deterministic switching of ferromagnetism at room temperature using an electric field. *Nature* **516**, 370–373 (2014).
13. Si, M. et al. A ferroelectric semiconductor field-effect transistor. *Nat. Electron.* **2**, 580–586 (2019).
14. Ievlev, A. et al. Intermittency, quasiperiodicity and chaos in probe-induced ferroelectric domain switching. *Nat. Phys.* **10**, 59–66 (2014).
15. Müller, K. A. & Burkard, H. $SrTiO_3$: an intrinsic quantum paraelectric below 4 K. *Phys. Rev. B* **19**, 3593 (1979).
16. Jerry, M. et al. Ferroelectric FET analog synapse for acceleration of deep neural network training. In *2017 IEEE Int. Electron Devices Meeting (IEDM)*, 6.2.1–6.1.4 (IEEE, 2017).
This work demonstrated multi-state (5-bit) weight cell/analogue synapse based on a ferroelectric field-effect transistor for deep neural network accelerator applications with a ×4 conductance modulation and ~75 ns program pulses.
17. Seo, M. et al. First demonstration of a logic-process compatible junctionless ferroelectric finfet synapse for neuromorphic applications. *IEEE Electron Device Lett.* **39**, 1445–1448 (2018).
18. Chung, W., Si, M. & Peide, D. Y. First demonstration of Ge ferroelectric nanowire FET as synaptic device for online learning in neural network with high number of conductance state and g_{max}/g_{min}. In *2018 IEEE International Electron Devices Meeting (IEDM)*, 15.2.1–15.2.4 (IEEE, 2018).
19. Ni, K. et al. SoC logic compatible multi-bit FeMFET weight cell for neuromorphic applications. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 13.2.1–13.2.4 (IEEE, 2018).
20. Sun, X., Wang, P., Ni, K., Datta, S. & Yu, S. Exploiting hybrid precision for training and inference: A 2T-1FeFET based analog synaptic weight cell. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 3.1.1–3.1.4 (IEEE, 2018).
21. Li, X. et al. Enabling energy-efficient nonvolatile computing with negative capacitance FET. *IEEE Trans. Electron Devices* **64**, 3452–3458 (2017).
22. Wang, Z. et al. Experimental demonstration of ferroelectric spiking neurons for unsupervised clustering. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 13.3.1–13.3.4 (IEEE, 2018).
This work experimentally demonstrated the concept of a ferroelectric field-effect transistor based oscillators and spiking neurons.
23. Fang, Y. et al. Neuro-mimetic dynamics of a ferroelectric FET-based spiking neuron. *IEEE Electron Device Lett.* **40**, 1213–1216 (2019).
24. Wang, Z., Khandelwal, S. & Khan, A. I. Ferroelectric oscillators and their coupled networks. *IEEE Electron Device Lett.* **38**, 1614–1617 (2017).
25. Fang, Y. et al. A swarm optimization solver based on ferroelectric spiking neural networks. *Front. Neurosci.* **13**, 855 (2019).
26. Oh, S., Hwang, H. & Yoo, I. Ferroelectric materials for neuromorphic computing. *APL Mater.* **7**, 091109 (2019).
27. Gokmen, T. & Vlasov, Y. Acceleration of deep neural network training with resistive cross-point devices: Design considerations. *Front. Neurosci.* **10**, 333 (2016).
28. Ni, K. et al. Ferroelectric ternary content-addressable memory for one-shot learning. *Nat. Electron.* **2**, 521–529 (2019).
29. Tan, A. J. et al. Experimental demonstration of a ferroelectric HfO_2-based content addressable memory cell. *IEEE Electron Device Lett.* (2019).
30. Wang, Y. et al. A 3 µs wake-up time nonvolatile processor based on ferroelectric flip-flops. In *2012 Proc. ESSCIRC* 149–152 (IEEE, 2012).
31. Mulaosmanovic, H., Mikolajick, T. & Slesazeck, S. Accumulative polarization reversal in nanoscale ferroelectric transistors. *ACS Appl. Mater. Interfaces* **10**, 23997–24002 (2018).
This work demonstrated non-trivial effects related to ferroelectric polarization switching at scaled lateral dimensions, namely meta-plasticity/polarization accumulation, single-domain-like switching and probabilistic switching, in short channel ferroelectric field-effect transistors.
32. Ni, K. et al. In-memory computing primitive for sensor data fusion in 28 nm HKMG FeFET technology. In *2018 IEEE International Electron Devices Meeting (IEDM)*, 16.1.1–16.1.4 (IEEE, 2018).
33. Mulaosmanovic, H., Mikolajick, T. & Slesazeck, S. Random number generation based on ferroelectric switching. *IEEE Electron Device Lett.* **39**, 135–138 (2017).
34. Luo, J. et al. Capacitor-less stochastic leaky-FeFET neuron of both excitatory and inhibitory connections for SNN with reduced hardware cost. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 6–4 (IEEE, 2019).
35. Florent, K. et al. Vertical ferroelectric HfO_2 FET based on 3-D NAND architecture: towards dense low-power memory. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 2.5.1–2.5.4 (IEEE, 2018).
36. Thirumala, S. K. & Gupta, S. K. Reconfigurable ferroelectric transistor—part I: device design and operation. *IEEE Trans. Electron Devices* **66**, 2771–2779 (2019).
37. Tasneem, N. & Khan, A. I. On the possibility of dynamically tuning and collapsing the ferroelectric hysteresis/memory window in an asymmetric DG MOS device: a path to a reconfigurable logic-memory device. In *2018 76th Device Research Conference (DRC)* 1–2 (IEEE, 2018).
38. Wang, Z. et al. Cryogenic characterization of a ferroelectric field-effect-transistor. *Appl. Phys. Lett.* **116**, 042902 (2020).
39. Chen, K.-Y., Tsai, Y.-S. & Wu, Y.-H. Ionizing radiation effect on memory characteristics for HfO_2-based ferroelectric field-effect transistors. *IEEE Electron Device Lett.* **40**, 1370–1373 (2019).
40. Shum, D. et al. Functionality demonstration of a high-density 1.1 V self-aligned split-gate NVM cell embedded into LP 40 nm CMOS for automotive and smart card applications. In *2015 IEEE Int. Memory Workshop (IMW)* https://doi.org/10.1109/IMW.2015.7150288 (IEEE, 2015).
41. Kanda, A. et al. A 24-MB embedded flash system based on 28-nm SG-MONOS Featuring 240-MHz read operations and robust over-the-air software update for automotive applications. *IEEE Solid-State Circ. Lett.* **2**, 273–276 (2019).
42. Tsuda, S. et al. First demonstration of FinFET split-gate MONOS for high-speed and highly-reliable embedded flash in 16/14 nm-node and beyond. In *2016 IEEE Int. Electron Devices Meeting (IEDM)*, 11.1.1–11.1.4 (IEEE, 2016).
43. Wei, L. et al. A 7Mb STT-MRAM in 22FFL FinFET technology with 4 ns read sensing time at 0.9 V using write-verify-write scheme and offset-cancellation sensing technique. In *2019 IEEE Int. Solid-State Circuits Conference-(ISSCC)*, 214–216 (IEEE, 2019).
44. Lee, K. et al. 1Gbit high density embedded STT-MRAM in 28nm FDSOI technology. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 2.2.1–2.2.4 (IEEE, 2019).
45. Jain, P. et al. A 3.6 Mb 10.1 Mb/mm2 embedded non-volatile ReRAM macro in 22 nm FinFET technology with adaptive forming/set/reset schemes yielding down to 0.5 V with sensing time of 5ns at 0.7 V. In *2019 IEEE Int. Solid-State Circuits Conference-(ISSCC)* 212–214 (IEEE, 2019).
46. Wu, J. et al. A 40 nm low-power logic compatible phase change memory technology. In *2018 IEEE Int. Electron Devices Meeting (IEDM)* 27–6 (IEEE, 2018).
47. Dünkel, S. et al. A FEFET based super-low-power ultra-fast embedded nvm technology for 22 nm FDSOI and beyond. In *2017 IEEE Int. Electron Devices Meeting (IEDM)* 19–7 (IEEE, 2017).
This work demonstrated scaled ferroelectric field-effect transistors in the fully depleted silicon-on-insulator (SOI) platform at the 22 nm node.
48. Trentzsch, M. et al. A 28 nm HKMG super low power embedded NVM technology based on ferroelectric FETs. In *2016 IEEE Int. Electron Devices Meeting (IEDM)*, 11–5 (IEEE, 2016).
49. Lyu, X., Si, M., Shrestha, P., Cheung, K. & Ye, P. First direct measurement of sub-nanosecond polarization switching in ferroelectric hafnium zirconium oxide. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 15–2 (IEEE, 2019).
50. Wei, Y. et al. Magneto-ionic control of spin polarization in multiferroic tunnel junctions. *npj Quant. Mater.* **4**, 1–6 (2019).
51. Cheema, S. S. et al. Enhanced ferroelectricity in ultrathin films grown directly on silicon. *Nature* **580**, 478–482 (2020).
This work demonstrated the thickness scalability of hafnium oxide based ferroelectrics.
52. Lederer, M. et al. Local crystallographic phase detection and texture mapping in ferroelectric Zr doped HfO_2 films by transmission-EBSD. *Appl. Phys. Lett.* **115**, 222902 (2019).

53. Grimley, E. D., Schenk, T., Mikolajick, T., Schroeder, U. & LeBeau, J. M. Atomic structure of domain and interphase boundaries in ferroelectric HfO_2. *Adv. Mater. Interf.* **5**, 1701258 (2018).
54. Ni, K., Chakraborty, W., Smith, J., Grisafe, B. & Datta, S. Fundamental understanding and control of device-to-device variation in deeply scaled ferroelectric fets. In *2019 Symposium on VLSI Technology*, T40–T41 (IEEE, 2019).
55. Maekawa, K. et al. Impact of homogeneously dispersed al nanoclusters by Si-monolayer insertion into $Hf_{0.5}Zr_{0.5}O_2$ film on FeFET memory array with tight threshold voltage distribution. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 15–4 (IEEE, 2019).
56. Migita, S., Morita, Y., Mizubayashi, W. & Ota, H. Preparation of epitaxial HfO_2 film (EOT=0.5 nm) on Si substrate using atomic-layer deposition of amorphous film and rapid thermal crystallization (RTC) in an abrupt temperature gradient. In *2010 Int. Electron Devices Meeting*, 11.5.1–11.5.4 (IEEE, 2010).
57. Chatterjee, K. et al. Self-aligned, gate last, FDSOI, ferroelectric gate memory device with 5.5-nm $Hf_{0.8}Zr_{0.2}O_2$, high endurance and breakdown recovery. *IEEE Electron Device Lett.* **38**, 1379–1382 (2017).
58. Chau, R. Process and packaging innovations for Moore's Law continuation and beyond. In *2019 Int. Electron Devices Meeting* 1.1 (IEEE, 2019).
59. Kozodaev, M. G. et al. Mitigating wakeup effect and improving endurance of ferroelectric HfO_2-ZrO_2 thin films by careful La-doping. *J. Appl. Phys.* **125**, 034101 (2019).
60. Muller, J. et al. High endurance strategies for hafnium oxide based ferroelectric field effect transistor. In *2016 16th Non-Volatile Memory Technology Symposium (NVMTS)* https://doi.org/10.1109/NVMTS.2016.7781517 (IEEE, 2016).
61. Ni, K. et al. Critical role of interlayer in $Hf_{0.5}Zr_{0.5}O_2$ ferroelectric FET nonvolatile memory performance. *IEEE Trans. Electron Devices* **65**, 2461–2469 (2018).
This work provides insights towards the design optimization of the gate stack of a modern ferroelectric field-effect transistor for improved reliability.
62. Yurchuk, E. et al. Origin of the endurance degradation in the novel HfO_2-based 1T ferroelectric non-volatile memories. In *2014 IEEE Int. Reliability Physics Symposium* 2E.5.1–2E.5.5 (IEEE, 2014).
63. Toprasertpong, K., Takenaka, M. & Takagi, S. Direct observation of interface charge behaviors in FeFET by quasi-static split CV and Hall techniques: Revealing FeFET operation. In *2019 IEEE Int. Electron Devices Meeting (IEDM)* 23.7.1–23.7.4 (IEEE, 2019).
This work demonstrated a method to calculate the number of trapped carriers at the interfaces that screen the ferroelectric polarization in a ferroelectric field-effect transistor.
64. Tan, A. et al. Hot electrons as the dominant source of degradation for sub-5nm HZO FeFETs. In *2020 Symposium on VLSI Technology* (IEEE, in the press).
65. Ando, T. Ultimate scaling of high-κ gate dielectrics: Higher-κ or interfacial layer scavenging? *Materials* **5**, 478–500 (2012).
66. Cheng, C.-H. & Chin, A. Low-leakage-current DRAM-like memory using a one-transistor ferroelectric MOSFET with a Hf-based gate dielectric. *IEEE Electron Device Lett.* **35**, 138–140 (2013).
67. Khanna, S. et al. An FRAM-based nonvolatile logic MCU SoC Exhibiting 100% digital state retention at VDD= 0 V Achieving Zero Leakage With < 400-ns Wakeup Time for ULP Applications. *IEEE J. Solid-State Circ.* **49**, 95–106 (2013).
68. Moise, T. et al. Demonstration of a 4 Mb, high density ferroelectric memory embedded within a 130 nm, 5 LM Cu/FSG logic process. In *Digest Int. Electron Devices Meeting* 535–538 (IEEE, 2002).
69. A century of ferroelectricity. *Nat. Mater.* **19**, 129 (2020).
70. Meterelliyoz, M. et al. 2nd generation embedded DRAM with 4X lower self refresh power in 22 nm Tri-Gate CMOS technology. In *2014 Symposium on VLSI Circuits Digest of Technical Papers* https://doi.org/10.1109/VLSIC.2014.6858415 (IEEE, 2014).
71. Lombardo, S. et al. Atomic-scale imaging of polarization switching in an (anti-)ferroelectric memory material: Zirconia (ZrO_2). In *2020 Symposium on VLSI Technology* (in the press).
This work studied the microstructural evolution during polarization switching in antiferroelectric zirconium oxide using in-situ transmission electron microscopy.

Acknowledgements

This work was supported by the Applications and Systems-Driven Center for Energy-Efficient Integrated Nano Technologies (ASCENT), one of six centers in the Joint University Microelectronics Program (JUMP), an SRC program sponsored by the Defense Advanced Research Program Agency (DARPA), the Center for Advanced Self-Powered Systems of Integrated Sensors and Technologies (ASSIST), an Engineering Research Center sponsored by the National Science Foundation (NSF), and the National Science Foundation (grant no. 1810005). We thank S. Yu, S. Mahapatra, W. van den Hoek, A. Raychowdhury, S. Salahuddin, K. Ni, S. Gupta, S.K. Thirumala, M.M. Islam and M. Hoffmann for insightful discussions.

Author contributions

All authors discussed the ideas and wrote the paper.

Competing interests

The authors declare that they have no competing interests.

Additional information

Correspondence should be addressed to A.I.K., A.K. or S.D.