1. Introduction

In recent years, artificial intelligence (AI) started influencing different aspects of our lives starting from smart sensing, facial recognition, image or pattern classification, natural language processing to autonomous vehicles, robotics, healthcare, and so on. In these applications, software-based AI performs the classification and pattern recognition tasks on digital CMOS computing platforms where data processing is clocked and takes place serially. This traditional form of digital computing is efficient for precision computing tasks but becomes inefficient and even unsustainable for data-intensive applications required for AI. Therefore, autonomous decision-making hardware needs to replace the software-based AI, which requires going beyond the traditional von Neumann computational architecture. [31]

Neuromorphic computing, inspired by neurobiological systems, forms one basis for development of such autonomous computational chips.[2–5] Application of neural networks in AI tasks requires a hardware system that is capable of operating at lower power and latency for both inference and training.[6] In the current computing framework, the synaptic weight values of the neural networks are stored in off-chip storages like the dynamic random access memories (DRAMs). In this framework, the efficiency of the networks, in terms of energy consumption and training time, get limited by the off-chip memory access bottleneck. As network sizes continue to expand, this is becoming a significant challenge for the computing hardware. Research in this direction confirmed that for a fully connected deep neural network (DNN), significant acceleration in training can be achieved by minimizing data movement between memory and processor by implementing on-chip storage and performing the computation and weight updates at the same node. This thrusts the development of crossbar compatible analog memory elements where the synaptic weight values are stored as the conductance of the memory components.[7,8]

In recent years, neuromorphic networks based on emerging memories have successfully demonstrated excellent performance in selected fundamental cognitive computing tasks, such as image, pattern and speech classification, reinforcement learning, etc.[9–11] Neural networks can be classified as artificial neural network (ANN) and spiking neural network (SNN). DNNs are the most common form of ANN, consisting of one input and one output layer of neurons together with one or more hidden layer of neurons fully connected to each other through the synapses. Deep learning requires electronic synapses to modulate their conductance in an analog manner to mimic the synaptic weight change during training by backpropagation.[12,13] In SNNs, a closer counterpart of the biological system, neurons...
communicate with each other through a sequence and timing of spikes, enabling them to process time-series data.\textsuperscript{[13]} In SNNs, the synapses between two firing neurons show long-term change in weights when the two connecting neurons fire relatively quickly. This long-term potentiation (LTP) or depression (LTD) of synaptic weights, i.e., increase or decrease in conductivity, respectively, emulate the spike-timing-dependent plasticity (STDP)\textsuperscript{[15,16]} that is used for supervised or unsupervised learning. Therefore, realization of neuromorphic computing hardware requires development of nanoelectronic devices that are able to emulate the functions of biological synapses and neurons.\textsuperscript{[17]} Emerging memory devices with analog conductance states, low-voltage operation, fast switching, and possibility of dense integration provides essential features needed for the successful implementation in neuromorphic chips. Recent years saw a significant surge in reports of both two-terminal (2T) and three-terminal (3T) memory devices that can emulate various synaptic functions.\textsuperscript{[18–20]} 2T memories such as oxide memristors, phase-change memories (PCMs), ferroelectric tunnel junctions (FTJs), and magnetic tunnel junctions (MTJs)\textsuperscript{[21,22]} and 3T devices including electrochemical transistors, ferroelectric field-effect transistors (FeFETs), and charge-trapping transistors showed promising performance as artificial synapses.\textsuperscript{[17,22–24]} Undoubtedly, the advantage of 2T devices arises from their reduced footprints, whereas the 3T devices provide a larger switching ratio, better control of analog conductance states, and improved cell selection within a crossbar array. While considering emulation of the neuronal functionalities for the neural networks, volatile memristors, PCMs, threshold switching devices, FTJs, and FeFETs have shown promising features.\textsuperscript{[25–29]} So far, the functions of biological synapses and neurons are successfully demonstrated in individual memory devices or small-scale proof-of-concept arrays; however, their large-scale implementations, needed for real-world applications, are still missing. Several combinations of CMOS with memristive devices have been intensely investigated over recent years.\textsuperscript{[31–33]} Broadly, these results indicate that for efficient online network training in hardware, in addition to abovementioned conditions, high linearity in synaptic weight update\textsuperscript{[31,34]} in response to programming pulses as well as endurance of $10^6$ cycles or more is essential for continuous set–reset operations of the synapses and integrate-and-fire neurons.\textsuperscript{[5]} Other essential features required are: 1) a wide resistance window for accurate continuous learning and multiple distinguishable states,\textsuperscript{[35]} 2) fast training and low-noise retention of conductance for maximizing the network throughput,\textsuperscript{[36]} 3) low programming and standby power\textsuperscript{[37]} to enable sensory information processing for long times even in limited-energy environments, and 4) integration possibility with CMOS processes.\textsuperscript{[31]} These criteria are in accordance with the International Roadmap for Devices and Systems for “Beyond CMOS” memory and information processing.\textsuperscript{[38]}

Since the first experimental demonstration of memristor concept, significant amount of work has been done to explore the potential of these devices in mimicking neurobiological sensory data processing tasks. Most of these efforts were concentrated on binary metal oxide 2T resistive random access memories (RRAMs) where conducting filament formation between the two electrodes under electric field determines the device on and off states. Also, electrochemical memories (CBRAM), PCMs, and spin-based memories, like MTJs, have been studied extensively.\textsuperscript{[17,24,39]} All different devices have certain aspects that are promising for energy-efficient in-memory-computing (IMC) architecture. However, most of these devices have a significant off current that makes them more power hungry, even while in standby mode and small resistive switching ratio ($R_{on/off}$) or eventual drift of conductance states are challenging. From the energy efficiency, fast operation, and stable data retention perspective, ferroelectric (FE) memory devices can offer unparalleled advantages over RRAMs and PCMs.\textsuperscript{[20,40,41]} Additionally, FeFETs can offer wide conductance window and high bit resolution, properties that are conducive to high-accuracy online learning. Although FE random access memories (FeRAMs), composed of one transistor one capacitor structures (1T1C), is an established technology, implementation of FE-based nonvolatile memories (NVMs) in embedded systems is impeded by the demanding process parameters like high temperature and need for certain lattice-matched substrates for epitaxial growth of the perovskite oxides. Also, maintaining ferroelectricity down to nanometer thickness scale in these conventional perovskite oxide FEs is often challenging that resulted in devices with thicker layer with a high operating voltage ($>10$ V). Additionally, large footprint of the capacitors is unsuitable for dense integration and destructive readout requires frequent programming. Ongoing development of high-quality, ultrathin FE films and possible low-voltage operation has revitalized the prospect of FE devices in NVM technology landscape in recent years.\textsuperscript{[42–44]} For FE-based neuromorphic systems, simulation based on ultralow current FTJ crossbar arrays predicted power efficiencies better than 150 Tera operations per second per watt (TOPS W$^{-1}$) for image classification using a trained ANN\textsuperscript{[37]} while in state-of-the-art oxide RRAMs energy efficiency of 187.62 Giga operations per second per watt (GOPS W$^{-1}$) is reported.\textsuperscript{[31]} The main reason for this enormous difference is that FE devices are voltage-driven instead of current-driven operations like RRAM components. FE components, therefore, offer exciting possibilities for integration into larger-scale memory and IMC arrays.

For application in electronic industry, an important requirement for a material of choice is the CMOS process compatibility. CMOS integration of electronic components and circuits allows monolithic fabrication of devices and circuits that produces a standalone chip, reducing the power loss at the interconnects, yielding more reliable interconnections, and helps minimize the parasitic losses by placing the device components in close proximity. In addition, integrated devices minimize lateral area requirement and unit device cost. As a CMOS compatible FE, doped HfO$_2$ has already gained considerable attention in recent years.\textsuperscript{[45,46]} However, the high crystallization temperature of the ultrathin films of most doped HfO$_2$ materials is still a challenge for their integration in post-CMOS processes. In addition, requirement of certain electrodes for proper crystallization of doped HfO$_2$ poses restriction on the choice of electrode materials. Also, limited endurance and endurance–retention compromise are issues that require further research. Another class of potentially CMOS postprocessing compatible FE materials is the FE polymers. Most studied FE polymers, in terms of memories, are polyvinylidene fluoride (PVDF) and its copolymer polyvinylidene fluoride-trifluoroethylene (PVDF-TrFE). These polymers have been used in post-CMOS processes for MEMS.
and other piezoresponse applications.\cite{47,48} However, their true potential for CMOS postprocessed FE memory components in neural networks has not been considered and explored yet. For heterogeneous CMOS integration or transferred freestanding device stacks on flexible substrates, perovskite FE devices based on BaTiO$_3$ (BTO), Pb$_{1-x}$Zr$_x$TiO$_3$ (PZT), LiNbO$_3$ (LNO) or SrBaTiO$_3$ (SBT), etc. can offer exciting properties.\cite{49} However, wafer scale, high-quality perovskite oxide FEs are still far from reality and their environmental impact is an issue of concern.

Another field of research, that is gaining tremendous interest in recent years, is smart and adaptive sensing for edge computing.\cite{32,50} Sensing technology has become integral part of our everyday lives with high emphasis on flexible, wearable, and bio-degradable technologies.\cite{51-53} With the onset of internet of everything (IOE), billions of internet-connected sensors are causing unprecedented amount of data flow to the cloud servers. Processing of this continuous flow of data is challenging and rather unsustainable for the current cloud-based computing platforms. Additionally, continuous data trafficking is a major source of large electrical power consumption that makes the edge devices unviable for operating at limited power environments. Data privacy and security is another additional concern.

Figure 1 shows the current computational schemes used for edge intelligence and the need for alternate approaches for energy and time-efficient, secure computing and sensing schemes. FE materials can play a big role in this data-driven industry by integrating both sensing and memory functionalities in the same device architecture.\cite{58,59} The current review gives an overview of these two research topics. After this introduction to the field of neuromorphic computing hardware and adaptive sensing, the article will discuss the state-of-the-art of various electronic and ionic devices with required functionalities, advantages and challenges of the FE components, restriction on choice of materials imposed by CMOS BEOL process parameters and flexible and wearable substrate compatibility, material physics of oxides and polymer FE, device physics of 2T FTJs and 3T FeFETs, application of oxide and polymer FTJs and FeFETs in NVM, synapses and neurons, application in DNN, SNNs, circuit integration issues, wafer scale homogeneity and CMOS back-end integration, and finally ending with a future perspective and outlook on the topics.

2. State-of-the-Art of Emerging Memories

In Table 1, some of the most promising emerging NVM technologies are compared in terms of the required performance matrices like $R_{on/off}$, switch speed, endurance, retention, write voltage, linearity in weight update, suitability for DNN or SNN applications, etc. with the mainstream CMOS memories. Some data are adapted from refs. \cite{17,60}. For an ideal forward-inference network, a large range of linearity and symmetrically programmable conductance states is preferred. This can ensure blind synaptic weight update during learning and can efficiently perform parallel vector matrix multiplication (VMM). Long data retention time is not always an essential condition, but is preferred in many applications, for instance, train-once inference-only applications require long-term data storage on-chip. High endurance and fast operation speeds are essential requirements for most computational tasks. Low power operation and reliability of the programmed states are always preferred. The

Figure 1. a) Schematic illustration of the data communication bottleneck faced by conventional computing schemes with the proposed idea of adding intelligence to the edge devices that will perform basic cognitive tasks already at the sensor node, significantly reducing the data overflow to the cloud servers. b) Computing-in-memory in memristive crossbar arrays where each cross section is a memory element capable of storing the information and perform VMM operation through change in conductance states by programmable ferroelectric polarization rotation. c) Capability of FE devices to do analog computation through controllable domain dynamics using cumulative nature of the programming pulses of different magnitude, duration, or interval. For a programming pulse of magnitude higher than the synaptic threshold voltage, a time invariance in output pulses is obtained, as each input pulse turns the device from off to on state, resulting in only postsynaptic amplitude modulation. However, when operated with subthreshold pulses, accumulation of multiple pulses can generate one output pulse, leading to both postsynaptic amplitude and frequency modulation.
reported results indicate that, with current state-of-the-art performance level, no single memory technology is capable of fulfilling all the requirements for IMC tasks. Therefore, based on the most critical functionality for a specific application, different emerging technology solutions can be considered. Nonvolatile nature of the FE polarization makes the FE memories one of the key material for NVMs. Also, the technology has the potential to reach some of the key performance indices (KPIs) indicated in the IRDS, especially the fast operation, high endurance, and possibility for dense integration. Additionally, analog conductance from multidomain FE devices and accumulative nature of switching makes them attractive for both DNN and SNN applications.

3. FE Memories

3.1. Advantages and Challenges of the FE Components

In addition to the above mentioned advantages of fast and energy-efficient operation of FE components, there are several other advantages that the FE memories possess that make them attractive. Major advantages include 1) forming free operation, 2) controllable intermediate states through gradual polarization reversal, 3) designing nonvolatile or volatile functionality in the same device through custom-engineered depolarizing field, 4) possibility of adding pressure and temperature sensing property in the memory element using piezo- and pyroelectric properties of the FE, 5) operation stability over a wide temperature range (from 4 K > Curie point of the FE), and under harsh radiation conditions. Other than these, 6) low-temperature processibility, structural flexibility, and low production cost of the polymer FE memories make them attractive for smart and wearable device applications. Additionally, 7) piezo- and triboelectric properties of FE materials can be utilized to generate power from human motion making them self-sustainable systems. However, there are some challenges that need to be overcome for realizing the full potential of these components. 1) Devices at nanoscale: Scaling down and dense integration of NVM components with FE materials converts the device operation from multidomain to single- or few-domain FE phase. Attaining and preserving the intermediate resistance states, arising from coexisting up and down polarized FE domains, is challenging to control in single- or few-domain limit. This affects the achievable number of analog synaptic weight states (bit resolution) leading to loss of classification accuracy. In the later sections of the article, it has been discussed how innovative material and interface engineering together with customized pulsing schemes could circumvent the problem. Also, for densely packed crossbars, FTJs are needed to be scaled drastically that results in low tunneling current and could subsequently increase read noise and readout time. One promising approach in this case could be the 3D vertical stacking of FTJs instead of their aggressive downsizing. 2) High spontaneous and remnant polarization at low voltage range: Most of the earlier studies reported on few hundreds of nm thick FE films where the threshold voltage ($V_{th}$) clearly exceeded tens of volts. For minimizing the size of the access transistors on the CMOS chip, it is essential to decrease the drive voltage to less than 3.3 V (for 180 nm technology node). In order to have a FE devices operating at less than 3.3 V or even less, clearly the thickness of the FE layer needs to be in the order of tens of nm or less where interface and pinhole effects can become significant. In order to avoid large gate leakage due to thin FE films, an additional dielectric can be useful, either as a separate layer or as a blend in the polymer FE film. Obviously, an important consideration in this case will be to maintain the effect of FE polarization as much as possible. 3) Energy-speed trade-off: The switching energy and time of the FE domains are inversely proportional and therefore choice of low voltage drive transistors in the CMOS circuit restricts the possibility of high-speed operation of the devices. A trade-off between the parameters

| Performance indices       | CMOS memories | Emerging memristive memories |
|---------------------------|---------------|------------------------------|
| On/off ratio              | 10⁴           | 10⁴                          |
| NOR Flash                 | 10⁴           | 10⁴                          |
| NAND Flash                | 10⁴           | 10⁴                          |
| Multilevel operation      | 2-bit         | 4-bit                        |
| Write voltage             | <10 V         | >10 V                        |
| Write energy (/bit)       | 100 pJ        | 10 fJ                        |
| Write time                | 1–10 μs       | 0.1–1 ms                     |
| Read time                 | 50 ns         | 10 μs                        |
| Endurance                 | 10⁹           | 10¹⁰                         |
| Retention                 | Long          | Medium                       |
| Drift                     | No            | No                           |
| Linearity                 | Low           | Low                          |
| Integration density       | High          | Very high                    |
| Suitability for DNN training | No             | No                           |
| Suitability for DNN inference | Yes           | Yes                          |
| Suitability for SNN applications | Yes           | Yes                          |

Table 1. Comparison of key performance matrices of the CMOS mainstream and emerging memory devices for their implementation in neuromorphic computing hardware. Adapted with permission. Copyright 2020, The Author(s). Published by MDPI, open access Creative Common CC BY license. More detailed data on 2T FTJs are in Table 2 and 3T FeFETs in Table 3.
is essential. Some feasible routes to overcome these challenges are discussed in the following sections of the review.

3.2. Materials for CMOS BEOL and Flexible Substrate Compatibility

Reported data on FE memories are mostly focused on either perovskite oxide FEIs, such as BTO, BiFeO₃, PbTiO₃, PZT, etc. or more recently on the doped HfO₂- based memories. Recent reports showed the promising behavior of polymer FE (PVDF-TrFE)-based electronic synapses and neurons. Operating at nanosecond timescales, [63–67] However, before converting them to a technologically matured solution, more research needs to be done, especially at the memory array level. For the doped-HfO₂ based FE memories, majority of reported materials and devices employed a high-temperature rapid thermal annealing to attain the noncentrosymmetric orthorhombic polar phase. [68–70] Although for a short time, this high-temperature annealing poses limitation on these materials for integrating in CMOS BEOL processes. However, recently a few studies found that it is possible to bring down the thermal annealing temperature of doped HfO₂ [46,71,72] or even making it annealing-free [73] without compromising their FE properties through implementation of proper growth processes or electrode materials. This review mainly focuses on those results with occasional mentions of high-temperature processes that is compatible with CMOS front-end-of-line (FEOL) or heterogeneous integration. It is important to mention here that perovskite FE devices using PZT, BTO, SBT, etc. have been used on flexible substrates as well, either in form of suspended nanoparticles in PVDF matrix or as transferred freestanding films. [49] These oxides, with their robust FE properties, are definitely promising for certain applications. However, wafer scale uniformity from these complex oxides has not been achieved yet. For flexible and wearable smart systems, P(VDF-TrFE)-based devices are still an excellent choice where under extreme bending and folding conditions devices can perform without any degradation or fatigue and are therefore promising for wearable smart systems and prosthetics. [74]

4. Origin of Ferroelectricity and Domain Dynamics

The origin of ferroelectricity in doped HfO₂- and PVDF-based copolymers has been discussed in detail in the literature. [46,75] Here, the polarization switching and relaxation mechanism together with the FE domain dynamics are discussed briefly for correlating the subsequent synaptic and neuronal functionalities of these devices with their transient switching dynamics. Ferroelectricity arises due to presence of a net electric polarization in certain class of dielectric materials. This net polarization can be controlled and tuned under application of an electric field of certain polarity. The net polarization in a FE material is a nonlinear function of electric field. When an electric field is applied on a FE material, above a certain field value, the material shows a sudden increase in its net polarization. This field is known as the coercive field (E_C). Above E_C, the material reaches its saturation polarization (P_s) value. When the applied field is withdrawn, the net polarization is maintained at a nonzero value, called the remnant polarization (P_r). On reversing the polarity of the applied field, the polarization direction rotates to the opposite direction and reaches the negative P_s value. P_r, P_s, and E_C are the characteristic parameters for a certain material. During the electric field cycling of a FE material, if the applied field is enough to saturate the polarization, the P–V hysteresis loop opening is broader and the loop is called a FE major loop. However, when the applied field is not enough to saturate the net polarization, the P–V hysteresis loop opening narrows down giving rise to a minor loop. The major and minor loop characteristics are important for designing stable analog operation of the FE memories and can be controlled utilizing a custom-programming scheme. [69] Also, the polarization retention is more stable when a FE major loop is obtained; therefore, a nonvolatile or a volatile memory performance can be expected based on the applied field strength to program a device. Minor loop operation is often associated with irreproducibility owing to randomness of microscopic domain switching and the “history effect.” [76] i.e., previous history of electric field cycling on the device, which can impact training accuracy of neural networks. Instead of continuous weight update, a full erase method before programming is able to recover the accuracy, however, at the cost of additional energy and latency. From material and device perspective, a dielectric capping layer for the FE, i.e., a linear voltage regulator in series with a nonlinear one, can improve the abruptness of polarization switching, thus providing better accuracy of analog states.

There are two main mechanisms involved in all FE devices that lead to performance degradation, i.e., imprint and fatigue. When the polarization hysteresis loop becomes asymmetric along the electric field axis, i.e., the +E_C and –E_C values are different for polarization switching, the FE components are known to have an imprint effect. Asymmetry at the FE–electrode interface and structural defects play a major role in imprint effect. Imprint can affect the symmetric weight update of the synapses using identical positive and negative pulses. Also, imprint effect is associated with a charge accumulation phenomenon that can eventually lead to a device breakdown. Fatigue in FE components is related to loss of P_s and P_r values over repeated switching cycles arising mainly due to charge trapping, leakage currents, or structural degradation due to bias stress. Memory endurance depends strongly on the FE fatigue properties.

Polarization switching in solid-state devices has been described by different domain-derived switching models. In the Kolmogorov–Avrami–Ishibashi (KAI) model, homogeneous formation of numerous nucleation centers followed by subsequent forward and sideways domain growth was proposed. The normalized change in polarization with time is expressed by the relation ΔP(t)/P₀ = 1 − exp [−(t/τ₀)^n], where t₀ is the characteristic switching time. The Avrami index, n, depends on the dimensionality of the domains and considered an integer taking a value of 3 for single crystals and 2 for epitaxial thin films. [77] The switching time follows the empirical Merz law, t₀ = t₀ exp(E/E₀), where E is the applied electric field, E₀ is the activation field proportional to the domain wall energy, and t₀ is the switching time at infinite applied field. KAI model was able to explain the polarization switching dynamics in FE single crystals and epitaxial thin films; however, in polycrystalline FE films, deviation from KAI model assumptions was observed where a distribution of switching time is observed experimentally. In the nucleation-limited switching (NLS) model, an alternative view of the switching dynamics is
proposed, where region-by-region nucleation and domain switching take place. Switching of the entire system in this case is governed by the domain nucleation statistics, rather than the domain expansion model, as proposed in the KAI model.

Depolarization field, $E_{\text{dep}}$, in a FE material arises due to an incomplete compensation of bound polarization charges by the layers adjacent to the FE layers. In FTJs, $E_{\text{dep}}$ inside the FE tunnel barrier can be estimated as, $E_{\text{dep}} = \frac{P}{C_0} \frac{\varepsilon_0}{\varepsilon_F} C_F + \frac{1}{C_0}$, where $\varepsilon_0$ is the vacuum permittivity, $P$ and $\varepsilon_F$ are the polarization and dielectric constant of the FE, and $\sigma_D$ is the screening charge density of the electrode. $E_{\text{dep}}$ can become significant in an ultrathin FE film where the screening length of the electrodes is high. For a FeFET, the expression for depolarization field can be approximated as, $E_{\text{dep}} = \frac{P}{C_0} \frac{\varepsilon_0}{\varepsilon_F} C_F + \frac{1}{C_0}$, where $C_F$ and $C_{IS}$ are the FE and series capacitance of the interface layer and semiconductor, respectively. From the equation, it becomes clear that for a FeFET, increase in the interface layer thickness and/or decrease in the FE layer thickness will result in higher $E_{\text{dep}}$. This will increase relaxation of the polarized domains and will lead to eventual loss in data retention. Based on these principles, $E_{\text{dep}}$ can be utilized in ANNs for custom-designed synaptic potentiation timescales or neuronal firing and recovery actions.

4.1. Doped HfO$_2$

Doped HfO$_2$ belongs to the class of incipient FEs where ferroelectricity can be induced and stabilized by different conditions such as mechanical confinement, doping, defects, stoichiometry variations, or exchange of (oxygen) isotopes. Doped HfO$_2$ films consist of tetragonal, orthorhombic, and monoclinic phases (Figure 2a). The orthorhombic crystalline phase content, responsible for the ferroelectricity in this material, can be induced by doping HfO$_2$ thin films with atoms of silicon, zirconium, or yttrium followed by an additional annealing step. The abovementioned stabilization conditions affect the thermodynamic free energy and can contribute to a FE phase formation. A theoretical description of this free energy modification is given by Landau–Ginzburg–Devonshire theory. Among the entire hafnia family, only in Hf$_0.5$Zr$_0.5$O$_2$ (HZO), <400 °C annealing temperature is reported to produce the

![Figure 2.](image1.png)
orthorhombic phases, but the crystallization temperature increases with decreasing thickness. In a recent report, Jiang et al. showed a cumulative heating approach can lead to FE phase in HZO under 400 °C, making them suitable for BEOL processes. In case of HZO thin films, theoretical calculations showed that a compressive stress along polar c-axis is necessary (in plane >8 GPa) for the monoclinic to orthorhombic phase transformation. In the polar orthorhombic phase, an applied electric field causes displacement of four O ions in the HfO2 unit cell inducing a large polarization (Figure 2b). Maximum value of P is reported in the literature is 53 μC cm⁻². Distribution and movement of oxygen vacancies under applied electric field play a major role in stabilizing the FE phase. For Al- and Si-doped HfO2, highest FE properties are seen at ≈9 mol% dopant concentration, while for HZO, a broader dopant range can show FE phase with highest FE properties reported for 50% doped HZO. Under repeated field cycling, the bulk of doped HfO2 undergoes a phase transformation from monoclinic to orthorhombic phase, and the interfaces show diminishing amount of nonuniform, defect-rich, tetragonal HfO2 layer. The evolution of these structuring aspects contributes to the increase in P, and the opening of the constricted P–V hysteresis that are known to occur with “wake-up.” The onset of the fatigue regime is correlated to an increasing concentration of bulk defects, which causes pinning of the domain walls.

4.1. Domain Dynamics in Doped HfO2

Dynamics of FE domain structures under an electrical bias has been studied in many doped hafnia systems. By a combination of piezoresponse force microscopy (PFM) and pulse switching techniques, it was shown that the nanoscopic polarization reversal mechanism and the wake-up process in these systems arise due to electrically induced domain depinning. Significant imprint in the doped HfO2 capacitors is caused by the different interface states that lead to asymmetric switching. It was found that decreasing annealing temperature results in higher amount of interface trap states leading to higher imprint and fatigue. Domain switching kinetics in doped hafnia system is generally described by NLS model characterized by a broad distribution of the local switching times. It is found that the domain velocity varies significantly throughout the switching process, indicating strong interaction with structural defects.

The switching times in HZO films of different thicknesses prepared under different O2 pressure were investigated by Ryu et al. A comparative study based on KAI and NLS switching models showed that NLS model can provide better fitting over the full range of polarization switching. The switching times could be modulated from 0.46 to 1.58 μs by varying the partial O2 pressure during deposition and film thickness. The activation field for polarization reversal increased with increasing O2 pressure with the pressure dependence being higher for thinner films. Also in Si-doped HfO2 based devices, the time, field, and temperature dependence of polarization reversal showed that when a domain is excited in the proximity of its coercive voltage, Vc, a probabilistic rather than a deterministic switching occurs. The nanoscale polarization kinetics of hafnia-based FEs reveals that the switching time can be exponentially reduced by increasing the applied bias for high-speed operations and by the correct tailoring of material parameters, like nucleation time and domain wall energy.

4.2. PVDF-Based Copolymers

In terms of molecular FE, P(VDF-TrFE) with chemical formula [(CH2–CF2)ₙ–(CF2–CF2)₁₉] is mainly discussed here as a prototypical member of the family, as it is the most reported system in modern flexible electronics. On the molecular level, FE polarization in P(VDF-TrFE) is a result of alignment of molecular dipoles formed by the difference in electronegativity between the electropositive hydrogen ions and electronegative fluorine ions attached to both sides of carbon atom. The molecular structure of P(VDF-TrFE), as shown in Figure 2c, consists of –CH2–CF2– and –CHF–CF2– units rigidly attached to the main-chain carbon atoms. The all-trans molecular conformation gives rise to molecular dipoles to be oriented perpendicular to the chain axis while the polarization reversal results from the rotation of these dipoles about the molecular chain. In the crystalline phase, molecules pack parallel to each other forming a quasihexagonal orthorhombic structure with a lattice spacing of ≈0.5 nm along the polar axis. Permanent electric dipole moments in P(VDF-TrFE), pointing from fluorine to hydrogen, gives rise to macroscopic polarization of P ≈ 10 μC cm⁻². In an applied electric field, the polarization reverses by rotation of the CH2 and CF2 units around the backbone of the polymer chain. Detailed microscopic investigation showed that the FE polarization switching in PVDF-based copolymer process involves a chain of processes starting from the nucleation, forward domain growth, and subsequent sideways growth. The timescale for each process is different resulting in a distribution of timescales for the observed domain switching. The nucleation and forward domain growth is typically very fast (in the range of 1 ps to 1 ns), whereas sideways domain growth could be comparatively slower (from several nanoseconds to seconds or even longer) depending on various intrinsic and extrinsic factors. Defects and structural disorders in the polymer introduced during synthesis can significantly influence the switching dynamics in P(VDF-TrFE) films by acting as nucleation centers and pinning sites for the domain wall motion, further escalating the complexity of the process. For precision computing, a sharp distribution of domain switching is preferable, and thus instead of continuous FE films, patterned nanostructures could be more promising. However, a larger distribution of switching times can be beneficial for probabilistic computing tasks and for hardware security through random number generation.

4.2.1. Domain Dynamics in P(VDF-TrFE)

Domain dynamics in PVDF-based copolymers reveal that depending on the film thickness and morphology, noticeable differences in polarization switching patterns can be obtained. PFM studies in P(VDF-TrFE) films showed at a relatively low thickness/size aspect ratio (<1/10), the average domain size w and film thickness d are related via Kittel relation, w = K dⁿ/₂ (with the constant K describing a balance between the domain wall and electrostatic energies). However, films with rice-like grains with high thickness/ grain aspect ratio (>1/5) were found to deviate significantly from this law. For films with high thickness/grain aspect ratio, the
surface energy effects contribute significantly leading to the formation of thermodynamically stable domain structures. Transformation of P(VDF-TrFE) films after annealing was found to be governed by a variation of the surface tension at the film-substrate interface. In the high aspect ratio films, an increase in the surface energy led to the formation of additional domain walls that is energetically unfavorable, thus resulting in larger domains. On the other hand, in lower aspect ratio films, the surface energy is significantly decreased leading to the formation of smaller anti-parallel 180° domains. Application of voltage pulses reverses the polarization in the grain. The size of the switched area depends strongly on the parameters of the switching pulse, with the smallest written domain being about 30 nm in size. It was found that PVDF films do not have any significant imprint behavior.\textsuperscript{[94]} The nanoscale features of the switching process include remote domain nucleation and spatially nonuniform wall velocity. Polarization reversal studies, performed with macroscopic dielectric measurements and by local probe techniques, revealed a rather complex switching behavior in the polymers that differs from both KAI and NLS models implying a strong dependence of the switching potential on the polymer defect structures. For a metal–FE-semiconductor (MFS) structure, the FE polarization reversal was shown to occur in close relation to charge compensation in the semiconductor layer.\textsuperscript{[95]} Application of a rectangular high voltage pulse on a n-Si/P(VDF-TrFE)/Au capacitor revealed an asymmetrical switching dynamics. Polarization switching on the positive bias side happens rapidly with the loss of the depletion layer due to fast accumulation of majority carriers while switching toward the negative side is markedly impeded by depletion layer formation. In a recent report, it is shown that due to different annealing temperatures and protocols, the morphology of the spin-coated ultrathin copolymer films are modified significantly, thus impacting the switching timescales and threshold fields of the FTJs. The best performing devices showed a projected switching timescale of subnanosecond range. An improvement in switching speed by seven orders of magnitude is achievable in these devices with an increase of the programming voltage by less than a factor of 2. Devices with lesser defects showed a good fit to NLS models over a wide field range where devices with higher amount of defects showed a fit to NLS model only at high field values.\textsuperscript{[66]} The mean domain switching times ($t_{\text{mean}}$), estimated from the NLS fit, when plotted against inverse field ($1/E$), gives a linear dependence (Merz law) and extrapolating the fit to infinite field value ($1/E = 0$), it is possible to calculate the theoretical limit of switching speed. For PVDF based systems, tens of picosecond of theoretical speed limit is calculated. This value is consistent with estimation from PFM results.\textsuperscript{[92]}

5. FTJs and FeFETs

5.1. Device Physics

5.1.1. FTJs

The rich physics of FE polarization switching, polarization retention, and relaxation in FTJ and FeFET architectures opens up an avenue for designing low-power, analog spiking neurons and synapses, similar to their biological counterparts. Figure 3 shows schematic illustrations of different device architectures and their working principles. In a traditional FeRAM, a FE capacitor stores the charge while a MOS transistor, connected in series, acts like a switch. Although robust in operation, high operating voltage and

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Ferroelectric memory cells consisting of a) one-transistor–one-capacitor (1T–1C) FeRAM cell where the nonvolatile data storage takes place in the capacitor and the connected MOSFET acts as a switch, b) one-transistor FeFET cell where the ferroelectric layer embedded in the FET gate stack acts as the charge storage medium, and c) an FTJ where change in FE polarization direction in the ultrathin tunnel barrier, sandwiched between two different metals, causes an asymmetric barrier potential resulting in different barrier heights for the tunneling electrons for the two polarization direction, causing resistive switching. d–f) The schematics of the corresponding device band diagrams.}
\end{figure}
large footprint of FeRAMs impeded their application in dense memory structures. In an FTJ, two different metal or semiconducting electrodes sandwich an ultrathin layer (1–5 nm) of FE barrier. In a varying electric field, the FE polarization switches between two perpendicular directions (up or down). Depending on the orientation of polarization in the tunnel barrier, charge carriers do either accumulate in or deplete from the interface layers of the electrodes to screen the bound polarization charges. For junctions with different electrodes, this screening effect produces an asymmetric barrier potential. As a result, polarization switching alters the mean barrier height for the tunneling electrons and, therefore, the electrical resistance of FTJs show bistable states with varying electric field. This phenomenon is known as the tunneling electroresistance (TER) effect. In general, two metal electrodes are used in FTJs. However, in recent times it is found that a semiconductor electrode at one end of the FE can give rise to larger resistive switching effect. This larger switching effect has its origin in the Schottky barrier formation at the FE/semiconductor interface when the polarization direction moves away from the interface. This Schottky barrier additionally changes the barrier width in addition to barrier height of the FTJs, leading to a large $R_{on/off}$ at room temperature.

Besides the crucial role of electrodes in the emergence of TER charge screening also affects the long-term stability of FE polarization due to aforementioned $E_{dep}$. Based on the FE polarization retention, a long- or short-term potentiation (LTP, STP) and depression (LTD, STD) can be designed in a device. Initially, most reports on FTJs concentrated on the bistable resistive switching effects; however, later on it is realized that FTJs can not only work as a binary memory, but multiple intermediate resistance states can be reliably accessed in FTJs through programmable control of partial switching of FE domains. In micron-scale devices, large number of FE domains are formed to maintain an energetically stable configuration and control of a mixed domain phase with coexisting up and down domains is comparatively easy. However, as devices are scaled down, and we approach from multi- to monodomain or a few-domain limit, precise control of intermediate states becomes a major concern. For nanoelectronic devices utilizing ultrathin FE films, another important aspect is the interface quality of the FE films with the metal or semiconductor electrodes. In oxide FTJs, in addition to the FE polarization reversal induced resistive switching, an electric-field-enhanced reversible migration of oxygen vacancies also plays a big role. Defects at the interface or inside the FE film can cause charge trapping sites that, over a significant cycle of operation, can build up significantly, hampering the normal device operation. In neural networks, this is a major concern for DNN training operation where $10^6$ cycles of on–off cycles are expected from the synaptic weight elements. Therefore, for any large-scale networks concerning the nanoscale FE devices, interface quality and control of FE domain dynamics, i.e., a combination of switching and relaxation timescale, are of critical importance. With proper design of material stack and configuration, an extremely energy-efficient and versatile neural network can be created.

### 5.1.2. FeFETs

In a FeFET, a FE material replaces the gate dielectric layer of a MOSFET structure. Therefore, an applied gate voltage polarizes the gate FE layer, making the net polarization pointing toward or away from the semiconductor channel. When the polarization direction points toward the semiconductor channel, electrons accumulate at the interface between the semiconductor and the FE layer in n-type semiconductors. This downward polarization, thus, results in increased conductivity in the n-type semiconductor channel, representing the device on state. Conversely, when the polarization direction points away from the semiconductor channel, a depletion of electrons occurs at the interface resulting in the decrease in drain–source current, leading to device off state. For a p-type semiconductor channel, similarly, a reverse FE polarization determines the device on and off state. In general, a gate voltage of more than 5 V magnitude is needed to fully rotate the polarization state in the 30–50 nm FE layer of the FeFET and thus opening the full memory window (MW). However, to keep the access transistors small in CMOS readout, the limitation comes in terms of maximum programming voltage. This requirement opens up the necessity for exploring novel materials interface and device design space for maximizing the memory performance within the 3.3 V operation limit. For analog memories, however, this opens up a possibility for attaining large number of intermediate states, where a relatively lower applied gate voltage results in partial polarization rotation of the FE, creating a mixture of up and down polarized domains. This coexisting up and down polarized domains neither fully accumulate or deplete the semiconductor channel resulting in intermediate conductance levels. However, partially polarized states are often not as stable as the fully polarized states. Additionally, a history dependence of the intermediate states is reported in the literature that makes designing of nonvolatile analog memory with the FeFETs challenging. Various FeFET devices are reported in the literature with different constituent semiconductors, FE, and device configurations. Here, some of the promising results are discussed for both hafnia- and PVDF-based FeFETs for their future prospect in CMOS integrated hardware neural network (HNN) and near or in-sensor computing.

### 5.2. Performance of FTJs as NVMs, Synapses, and Neurons

#### 5.2.1. FTJs as NVMs

Solid-state FTJs, investigated within the last decade, are mainly based on oxide perovskite FEs. These devices have shown promises for their application in nonvolatile resistive memory elements. Large $R_{on/off}$ reproducibility, high-speed, and low operation energy are the major advantages of FTJs compared to the FE capacitor-based FeRAMs. From technology perspective, endurance of FTJs needs special attention, which is currently limited to the order of $10^6$, mostly. A very high endurance value ($>10^{14}$ cycles) for FeRAMs based on thick FE films is reported. This indicates that in ultrathin films, a right combination of FE-electrode materials and their carefully engineered interfaces must be found to reach this KPI. For oxide FE, defects such as oxygen vacancies in the tunnel barrier are a critical parameter controlling endurance, as repeated operation leads to trapped charges in the barrier, accelerating the breakdown of the device upon frequent cycling. The retention of the FTJs is also critical for NVM applications. Another important factor is the scalability of FTJs. In FTJs, scalability is not restricted
by its readout process like the classical FeRAMs. For oxide perovskite and polymeric FTJs, devices with dimensions of the order of 200 nm have been demonstrated. Results with devices <100 nm need further investigations for evaluating the scalability limit of FTJs and assess their potential as NVMs and as memristors in massively parallel architectures. In Section 6.5, low readout current of the ultrascaled FTJs and their possible solutions have been discussed.

Table 2 summarizes SoA KPIs of 2T memories based on HZO and P(VDF-TrFE). In some cases, FE diode and capacitor data are used as FTJ data are still rather limited. Both HfO2 and P(VDF-TrFE) memories show potential for long-term data retention. However, in many cases, a thin insulating oxide layer is unavoidable at the FE–semiconductor interface to improve the interface quality and reduce trapping sites. This insulating layer has an adverse effect on the data retention. Thus, optimal performance requires an application-specific, suitable architecture (MFM or MFIM). As can be seen from Table 2, many of the reported structures can offer suitable Ret (cycles) operation in low-voltage range, significant MW, fast operation speed, and low thermal budget fabrication, suitable for BEOL processes. Multibit analog operation in many cases has not been studied; however, their switching pattern indicates that it is possible to have controllable multibit operation from these devices. In many cases, the endurance measurements were not carried out beyond 10^3 cycles. While retention properties are easily extrapolated, endurance data need to be studied experimentally to verify their true potential for training architectures. The best reported data for HZO-based 2T memories exceed 10^6 cycles of operation, while for P(VDF-TrFE)-based capacitors prevention of delamination of the top electrode is shown to improve the endurance significantly. In P(VDF-TrFE), electron-induced phase decomposition due to repeated electric field cycling can cause accumulation of gases, leading to device fatigue and eventual failure. When the gas barrier is removed through use of a transparent electrode, a virtually fatigue-free FE capacitor is realized. The best value in P(VDF-TrFE) capacitor with PEDOT:PSS electrode is >10^8 cycles, which is suitable for DNN training applications.

### 5.2.2. FTJs as Synapses

Besides their use as binary NVM, FTJs can be used as memristive devices for artificial synapses in neuromorphic architectures. These FE memristors interconnected in HNNs are able to demonstrate unsupervised learning with STDP. Ultralow power HZO-based FTJ synapses have been demonstrated for HNN applications. For the networks, high-density 3D vertical HZO FTJs were fabricated that showed a P value of ∼16 μC cm^-2 and nonvolatile storage at 100 ns for over 10^3 cycles. Biological synaptics like LTP, LTD, and symmetric STDP at an energy consumption of 1.8 pJ per spike were demonstrated (Figure 4).

A pattern training in a 6 × 6 FTJ array showed a strong tolerance to the input faults and variations. The simulation of pattern classification and recognition, performed using the backpropagation algorithm, showed >96% accuracy providing a high-potential implementation of the HZO-based FTJs in ANNs. Annealing temperature of 450°C makes them suitable for CMOS post-processing compatible; however, an operating voltage >5 V, for operating in major loop, can still be a concern for the low-voltage CMOS readout. Further optimization of stack structures can be proved beneficial in this regard. Other reports on HZO-based FTJ synapses also confirmed efficient STDP-based learning.

In another work, FTJs based on metal/Al2O3/HZO/Si double-barrier structure was demonstrated to perform as an electronic synapse. The interfacial Al2O3 layer and silicon substrate were shown to enable sizable TER, even when the thickness

| Device architectures | On/off ratio | Multibit operation | Write voltage | Write time | Retention | Endurance (cycles) | MW [V] | Process temperature [°C] | References |
|----------------------|-------------|--------------------|---------------|------------|-----------|--------------------|-------|--------------------------|------------|
| TiN/HZO/Pt           | 15          | –                  | 3 V           | –          | >3 × 10^3 s | >10^3              | 4 V   | 425                      | [161]      |
| LSMO*/HZO/Pt         | 10^3        | –                  | 5 V           | –          | –         | 6 V                | 700   | –                        | [162]      |
| p-Si/SiO2/HZO/W      | 30          | –                  | 2 V           | –          | >8 × 10^4 s | –                  | 2.5 V | 500                      | [43]       |
| TiN/HZO/TiN          | 10^3        | –                  | 6 V           | <20 ns     | >5 × 10^6 s | >10^1              | 2 V   | 400                      | [116]      |
| n-Si/SiO2/HZO/TiN    | 30          | Yes                | 3 V           | –          | >10^3 s    | <10^3              | 6 V   | 500                      | [163]      |
| TiN/HZO/Mo (cap)     | –           | –                  | 2 V           | –          | >10^4 s    | >10^4              | 3 V   | 500                      | [42]       |
| Pt/HZO/TiN           | 20          | Yes                | 5 V           | <100 ns    | –         | >10^3              | 7 V   | 450                      | [118]      |
| W/P(VDF-TrFE)/Au     | 10          | –                  | 1.5 V         | –          | –         | 3 V                | 25    | [113]                    |
| Al/P(VDF-TrFE)/PTFE/Al| 10^3        | –                  | 4 V           | –          | >10^3 s    | –                  | 6 V   | 170                      | [164]      |
| ITO/P(VDF-TrFE)/Au   | >10^3       | Yes                | <3 V          | –          | <10^3 s    | >10^3              | 2 V   | 140                      | [102]      |
| NISTO*/P(VDF-TrFE)/Au| >10^3       | Yes                | <5 V          | <20 ns     | >10^4 s    | >10^1              | 3 V   | 140                      | [102]      |
| Au/P(VDF-TrFE)/ZnONPs/n++Si | >10^3 | –                  | <4 V          | –          | >10^4 s    | >10^4              | 6.5 V | 140                      | [165]      |
| Mo/MoO3/FBT-P(VDF-TrFE)/Al | >10^3 | Yes                | 15 V          | 10 ms      | >10^3 s    | >10^4              | >6 V  | 135                      | [166]      |
| Au/P(VDF-TrFE)/PEDOT:PSS (Cap) | –      | –                  | 40 V          | –          | –         | >10^6              | 140   | [117]                    |
| ITO/P(VDF-TrFE)/PEDOT:PSS (Cap) | –      | –                  | 15 V          | –          | –         | >10^6              | 20 V  | 140                      | [167]      |

a) At least one component is not BEOL compatible.
of HZO is above 10 nm. The Al2O3/HZO synapses show symmetric potentiation/depression characteristics and widely tunable conductance under constant modulation of pulse amplitude. However, under constant amplitude and pulse width, the potentiation and depression were highly asymmetric. Also, the operating voltage of the devices was >7 V that needs further reduction for low-voltage and low-power applications.

Emulation of key synaptic functions was demonstrated using 2T FTJ memristors with a spin-coated P(VDF-TrFE) (70:30) tunnel barrier sandwiched between Au- and Nb-doped SrTiO3 (NSTO) electrodes (Figure 5). The FTJs exhibited analog switching behavior over a five-order-of-magnitude range. The semiconducting bottom electrode-induced Schottky barrier formation led to this large on-off ratio and a rectifying I–V characteristics that can potentially allow selector-less integration of these FTJs in crossbar arrays. The FTJs reproducibly mimicked LTP/STP, LTD/STD, paired-pulse facilitation and depression (PPF/PPD) and STDP on timescales down to nanoseconds. A good polarization retention, switching cycle reproducibility, and large range of accessible conductance states provide large versatility for these FTJs in the design of HNNs. Besides, operation at nanosecond time scale opens an avenue toward the realization of ultraenergy-efficient electronic synapses. Estimation of energy consumption in these synapses indicates that potentiation and depression of the FTJs require an energy of 1.1 pJ and 170 fJ, respectively. While this is already comparable to the synaptic energy usage in the brain, possibility of further reduction of energy consumption is predicted with downscaling the junction sizes. Assuming linear variation of tunneling current with junction area, an energy consumption of $1.2 \times 10^{-17}$ J/μm² for potentiation and $1.9 \times 10^{-18}$ J/μm² depression was predicted.

In a recent work, a significant improvement in conductance update linearity is reported from the similar architecture with the nonlinearity factor decreasing from 2.7 to 1.2 for potentiation and from 5.2 to 1.5 for depression due to application of nanosecond pulsing schemes and higher domain pinning sites (Figure 5g). Here, it is important to point out that many of this exciting properties arise due to FE/NSTO interface. NSTO is an n-doped semiconductor where large band bending occurs due to FE polarization induced field. However, NSTO is neither a CMOS postprocessing compatible nor a material suitable for integration on flexible substrates. Therefore, other low-temperature processable semiconductors need to be tested in similar FTJ structures. Oxide semiconductors such as indium oxide (In2O3), indium gallium zinc oxide (IGZO), or 2D semiconductors like MoS2 or WSe2 can be promising alternatives in this respect.

5.2.3. FTJs as Neurons

For neuromorphic architectures, leaky-integrate-and-fire (LIF) model of neurons is widely adopted due to their relatively simple mathematical description, yet providing sufficient accuracy in replicating essential biological features. The neuronal dynamics can be characterized in the following major features: integration of weighted synaptic inputs arriving from other neurons,
subsequent firing when a certain threshold is reached and the recovery to their initial states. These momentary firing dynamics can be reproduced in FE devices using transient switching characteristics of the FE materials.

In general, in FTJs and FeFETs, the retention of the programmed conductance states is stable due to nonvolatile nature of FE polarization. However, absence of proper screening of the polarization-bound charges can lead to significant $E_{\text{dep}}$ causing relaxation of the programmed conductance states. Therefore, through proper control of relaxation dynamics of FE polarization, it is possible to design a LIF neuron. Two easy ways to control the LIF behavior in an FTJ involves control of relaxation by modulating the screening charge-carrier density of the electrodes and by changing the FE film thickness. In a recent work, it is shown that by decreasing the charge carrier density of the semiconducting electrode, a crossover from a NVM- to a LIF-type neuronal behavior can be obtained (Figure 6).

LIF response was recorded in an FTJ with P(VDF-TrFE) as the FE and NSTO as the semiconducting bottom electrode, with 0.25 wt% Nb doping. The tunneling current of the FTJ showed oscillatory response when subthreshold pulses of magnitude $+0.7\, \text{V}$ and duration of $5\, \text{ms}$ are applied on the device at a time interval of 100 ms (Figure 6a–c). When the initial state of the FTJ is set close to the off state, the current output resulted in bundles of spikes with a width of $\approx 300\, \text{ms}$ and a frequency of 2.2 Hz. The $+0.7\, \text{V}$ pulses are below threshold for off to on transition, and consequently, the system needed to integrate the effect of a series of pulses before triggering a response. With increasing amplitude of input pulses, the width of the current pulse bundles increased and their frequency decreased. Also, the duration of the individual voltage pulses was found to determine the width and frequency of the current pulse bundles. For a pulse amplitude of $+1.0\, \text{V}$, i.e., when the input voltage is close to threshold, a series of 5 ms pulses with an interval of 100 ms no longer generated an oscillating output current. A more pronounced integrate-and-fire response is reported from the same device with the application of longer voltage pulses. An example is depicted in Figure 6e–g where voltage pulses of amplitude $+1.0\, \text{V}$ and a duration of 200 ms and interval of 800 ms produced a sudden integrate and fire response after about 30 s. Although the tunneling current shows initial oscillatory response, it suddenly increases, mimicking the integrate-and-fire behavior of a neuron. The first firing event marks threshold switching from a nearly off state to an on state. After this first firing event, the current amplitude decreases gradually to an intermediate level. The integration time required for LIF switching can be programmed by modulation of amplitude and frequency of the voltage pulse train where a frequency of 2.3 Hz reduced the integration time to 14 s. Further decrease in the integration time to 5.5 s with voltage pulses of amplitude $+1.1\, \text{V}$ led to continuous firing but at the expense of neuronal recovery. This is an indication that besides control over the integration time, the pulse parameters also determine the relaxation of the tunneling current after the firing event with stronger and more frequent pulsing prolongs the FE polarization relaxation time. Because firing and recovery in the FTJ-based neuron is a sensitive balance between FE polarization switching and retention, $E_{\text{dep}}$ of FTJs also needs careful designing for a proper LIF functionality with most important parameters being the charge carrier concentration in the semiconducting electrode and the thickness of the FE tunnel barrier. This finding opens a new

Figure 5. The organic FTJ memristor. a) Schematic illustration of the FTJ sample. In transport measurements, we grounded the Nb-doped SrTiO$_3$ (NSTO) substrate and applied voltage pulses to the Au top electrode (except for STDP). b) PFM phase images of a P(VDF-TrFE) film area that is subsequently written by applying $-5$, $+2$, $-2$, and $-5\, \text{V}$ to the scanning probe tip. The blue arrows illustrate the net perpendicular polarization. c) Interface band alignment for saturated (large arrows) and intermediate (small arrows) polarization states. If the polarization points away from the Nb-doped SrTiO$_3$ electrode (lower panels), electrons in the n-type semiconductor deplete from the interface. In this case, positively charged immobile donor ions (blue dots) screen the bound polarization charges in P(VDF-TrFE) (minus signs), causing the formation of a Schottky barrier. Upon polarization reversal, electrons accumulate at the Nb-doped SrTiO$_3$ interface (red dots), which eliminates the Schottky barrier (upper panels). The FTJ potentiation and depression under (d) gradually increasing and (e) identical voltage pulse trains. f) STDP learning response and g) improvement in weight update linearity with 20 ns pulse programming compared to 100 ns pulses. a–f) Reproduced with permission. Copyright 2019, The Author(s). Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Creative Commons CC-BY-NC-ND license. g) Reproduced with permission. Copyright 2021, The Author(s). Published by Royal Society of Chemistry, Creative Commons Attribution (CC BY) license.
possibility of designing the entire neural network using similar kind of FTJs only by designing a desired doping profile in the semiconductor bottom electrode. Future research in this direction would be interesting to explore the effect of different CMOS BEOL compatible semiconductors and FE.

5.3. Performance of FeFETs as NVM, Synapses, and Neurons

5.3.1. FeFETs as NVMs

Table 3 summarizes some of the KPIs of the BEOL compatible FeFET components. In comparison to the 2T FTJs, FeFETs have larger $R_{on/off}$ and better control over the intermediate resistance states. However, one major challenge of the hafnia-based FeFET technology is reported to be degradation of endurance and variability of switching voltage for ultrascaled devices. One of the recent works by Zhou et al. used targeted programming schemes that showed that with increased delay times after both program and erase pulses, both lower and higher $V_{th}$ shifts can be reduced in FeFETs. Additional increase in the pulse amplitude to $+1.1$ V lowers the integration time to 5.5 s. Reproduced with permission. Copyright 2019, The Author(s). Published by AIP Publishing. Creative Commons Attribution (CC BY) license.

Figure 6. (Top) Oscillatory output current of a polymeric FTJ with a NSTO electrode of Nb doping concentration of 0.25 wt%. The junction is operated near the off state by continuous pulsing with 100 ms intervals. The voltage pulse amplitude and duration are a) $+0.7$ V and 5 ms, b) $+0.8$ V and 5 ms, and c) $+0.8$ V and 10 ms. d) Frequency of the current pulse bundles as a function of voltage pulse amplitude for 5 and 10 ms pulses. (Bottom) Emulation of integrate-and-fire behavior using the same FTJ. The junction is initialized near the off state. e) Continuous pulsing with 200 ms $+1.0$ V pulses at a frequency of 1 Hz produces a sudden increase in the tunneling current after 30 s. f) The integration time before the first firing event decreases to 14 s if the pulse frequency is set to 2.3 Hz. g) An additional increase in the pulse amplitude to $+1.1$ V lowers the integration time to 5.5 s. Reproduced with permission. Copyright 2019, The Author(s). Published by AIP Publishing. Creative Commons Attribution (CC BY) license.

From a wide range of literature data, it can be safely conjectured that channel semiconductor material, optimal gate stack, and geometry design can play a major role in determining the FeFET properties. High $R_{on/off}$ is a general property of most reported structures; however, switching voltage, analog states,
retention, and endurance depend critically on the abovementioned properties. For instance, Park et al.\textsuperscript{[123]} reported a large switching response with a counterclockwise hysteresis in FeFETs with a 100 nm P(VDF-TrFE) as the gate FE and Si as the channel semiconductor. A 1.3 V MW was reported in the $I_D-V_G$ loop when gate voltage was swept between −3 and 5 V with $V_D$ of 0.5 V. However, integration of Si as the channel semiconductor in the BEOL processes is not feasible and hence low thermal budget materials like the oxide and organic semiconductors provide viable options. Figure 7 shows examples of oxide and one organic semiconductor based FeFET operations. Yoon et al.\textsuperscript{[124]} reported one of the most promising structures with P(VDF-TrFE) as the gate insulator and oxide semiconductor Al−Zn−Sn−O layers, in which thin Al$_2$O$_3$ is introduced between two layers, as the channel material. The field-effect mobility, subthreshold swing, on/off ratio, and gate leakage currents were obtained to be 32.2 cm$^2$ V$^{-1}$ s$^{-1}$, 0.45 V decade$^{-1}$, 10$^8$, and 10$^{-11}$ A, respectively. The MW of the FeFET was 7.5 V with a gate voltage sweep of −10 to 10 V, and remained at 1.8 V, even with a lower voltage sweep of −6 to 6 V. The transmittance of the fabricated device was more than 90% at the wavelength of 550 nm that makes it promising in transparent surface applications. Xu et al.\textsuperscript{[144]} showed two ultrathin AlO$_x$ interfacial layers sandwiching an ultrathin FE film with a low $E_c$ can decrease the FeFET operation voltage significantly. Downscaling thickness of the FE film does not affect the gate leakage due to insertion of two ultrathin AlO$_x$ layers at both sides of the FE film. The endurance, mobility, and retention were found satisfactory, both in stable and under bent conditions.

A transparent and flexible FeFETs on a polyethylene-naphthalate (PEN) substrate was demonstrated by Lee et al.\textsuperscript{[123]} using amorphous indium gallium zinc oxide (a-IGZO) as the channel semiconductor and P(VDF-TrFE) as gate insulator. High-performance transistors enduring hard bending tests confirmed P(VDF-TrFE) phase formation process do not damage the underlying a-IGZO channel. This low-temperature, low-damaging process of P(VDF-TrFE) is one of the most suitable features for the oxide FeFETs. With the development of a reliable process, a non-volatile pixel memory for flexible displays could become a very attractive application for the polymeric FeFETs. Similar features are reported with ZnO as the oxide channels; however in this case, MFIS structure was employed as the P(VDF-TrFE) formation process degraded the crystalline quality of ZnO.\textsuperscript{[126, 127]}

\textbf{Effect of Ambience:} Ferroelectric oxides or fluoropolymer PVDF and its copolymers are generally insensitive to humidity, oxygen, or thermal changes in normal atmospheric condition limits. However, FeFETs involve a semiconductor channel which might have a strong or weak sensitivity to ambient conditions. For instance, 2D semiconductors or organic semiconductors are very sensitive to humidity and oxygen: while oxide semiconductors can show eventual drift of conductance over time. To protect these devices from fast or gradual degradation, adoption of a proper encapsulation strategy is, therefore, essential.

\subsection*{5.3.2. FeFETs as Synapses}

Doped hafnia-based solid state synapse is demonstrated using a single FeFET integrated in a 28 nm HKMG technology.\textsuperscript{[128]} The nonvolatile, gradual switching of the Fe hafnia is exploited in these devices to continuously tune the conductivity of the transistor channel. This gave rise to continuous synaptic weight

---

**Table 3.** Key performance matrices of ferroelectric three-terminal memory devices for their implementation in neuromorphic computing hardware. First half of the table shows state-of-the-art numbers from BEOL compatible structures where the latter half shows performance matrices of materials with possibilities of BEOL integration.

| Device architectures | On/off ratio | Multibit operation | Write voltage | Write time | Retention endurance | Endurance (cycles) | MW [V] | Process temperature [°C] | References |
|----------------------|--------------|--------------------|---------------|------------|---------------------|-------------------|--------|---------------------------|------------|
| W(G)/HZO/WO/Pd(D-S)  | 10$^4$       | Yes                | 3 V           | <100 ns    | >10$^4$            | 10$^4$            | 0.45 V | 350                       | [131]      |
| TaN/HZO/ZrO/SiO/Si  | 10$^4$       |                   | 3 V           | <3 V       | >10$^4$            | 10$^4$            | 3 V    | 550                       | [168]      |
| W(G)/HZO/HfO/Si/Si  | 10$^6$       |                   | 3 V           | <100 ns    | >10$^4$            | 10$^4$            | 1 V    | 500                       | [169]      |
| TiN(G)/HZO/WO/W(D-S)| 4            | Yes                | 4 V           | >10$^3$    |         | >10$^3$            | 4 V    |                           | [130]      |
| TaN(G)/HZO/IGZO/TiAl(D-S) | 10$^4$     |                   | 2 V           | <1 μs      | >10$^4$            | 10$^4$            | 1.6 V  | 500                       | [171]      |
| Au(G)/TiN/HZO/HfO/MoS$_2$/Ni(D-S) | 10$^8$ | Yes                | 3 V           | <10$^4$    |         | >10$^4$            | 4 V    | 400                       | [172]      |
| TiN(G)/HZO/IGZO/AI(D-S) | 10$^4$     |                   | 5 V           | >10$^4$    |         | >10$^4$            | 2 V    | 400                       | [18]       |
| Al(G)/P(VDF-TrFE)/MoS$_2$/Cr/Au(D-S) | 10$^8$ |                   | 5 V           | <10$^4$    | >10$^4$            | 1 V    | 200                       | [173]      |
| Au(G)/P(VDF-TrFE)/PTAA/AI | 10$^6$ |                   | 20 V          | >10$^4$    |         | >10$^4$            | 1 V    | 140                       | [174]      |
| Au(G)/P(VDF-TrFE)/MoS$_2$/Gr(D-S) | 10$^7$ |                   | 15 V          | >600 s     | >15 V             | 15 V   | 140                       | [175]      |
| Au(G)/P(VDF-TrFE)/Si/Au(D-S) | 10$^8$ |                   | 3 V           | >10$^4$    |         | >10$^4$            | 1 V    | 165                       | [123]      |
| Au(G)/P(VDF-TrFE)/PV/Al(D-S) | >10$^4$ |                   | 15 V          | <1 ms      | >10$^4$            | 1 V    | 20 V                       | [176]      |
| Al(G)/P(VDF-TrFE)/MEH-PP/Al(D-S) | >10$^4$ |                   | >40 V         | <1 ms      | >10$^4$            | >60 V  | 140                       | [177]      |
| Au(G)/P(VDF-TrFE)/In$_2$O$_3$/NW/Al(D-S) | 10$^8$ |                   | 10 V          |           | >10$^4$            | 2 V    | 130                       | [178]      |
| Al(G)/P(VDF-TrFE)/MoS$_2$/Cr(Au(D-S)) | 10$^6$ |                   | 3 V           | >10$^4$    |         | >10$^4$            | >8 V   | 200                       | [147]      |
| Al(G)/P(VDF-TrFE)/Al$_2$O$_3$/ZnO/ITO(D-S) | 10$^8$ |                   | 5 V           | >10$^4$    |         | >10$^4$            | 3 V    |                           | [127]      |

\textsuperscript{a)At least one component is not BEOL compatible.}
By adjusting the channel FE thickness from 15 to 8 nm, retention tests performed over 4-bit depth showed low noise writing energy of fJ range and fast programming time of 40 ns. A schematic configuration of the flexible FeFET NVM revealed a 103 terms of area, latency, energy, and accuracy of a FeFET synaptic architecture like the two transistor-one FeFET (2T1F) hybrid weight cell. As discussed earlier, in FE analog memories, bit precision starts to become limited in scaled devices due to reducing number of available FE domains and inherent stochasticity in the FE domain switching behavior. In general, implementation of off-chip training followed by reduced bit precision for inference mode and on-chip learning and inference with reduced bit precision have been proposed. However, both strategies demand efficient training algorithm taking into consideration the quantization in synaptic weights.

Analog response from FeFET synapse is reported in many other works using voltage-dependent partial polarization switching in multidomain FE thin films. Recent experimental works have shown the ability to program FeFETs with voltage pulse widths as low as 50 ns while the programming voltage as low as 1.8 V is reported through proper engineering of the gate stack by adding an additional metal layer between the FE capacitor and MOS capacitor. In these devices, 32 conductance states (equivalent to 5-bit precision) are reported for 60 μm2 FeFET devices while 8 nonoverlapping conductance states (equivalent to 3 bits) were achieved in a 0.25 μm2 size devices. The precision of synaptic weight was proposed to be further improved by change in architecture like the two transistor-one FeFET (2T1F) hybrid weight cell that allows up to 64 states with improved nonlinearity and asymmetry factors. Ali et al. reported high endurance in hafnia-based FeFETs without retention penalty and analog states by implementing lamination of the FE layers.
Significant amount of work on PVDF-based FeFETs showed potential of these devices in mimicking neuroplasticity. In a recent work, Tian et al.\cite{63} reported that the conductance of P(VDF-TrFE)-based FeFETs, with 2D semiconductor MoS2 as channel material, can be precisely manipulated to vary between more than 1000 intermediate states with the highest switching ratio of \(10^4\). This continuous resistive switching in the MoS2 channel, resulting from the controllable FE domain dynamics, mimicked the synaptic plasticity behavior such as LTP and LTD and STDP. In addition, the device showed no sign of fatigue or failure even after 10\(^7\) cycles of continuous operation. Together with an ultralow energy consumption for each synaptic operation (less than 1 fJ), these devices can be considered as one of the potential contenders for the massive neural architecture in bioinspired networks.

5.3.3. FeFETs as Neurons

Accumulative nature of polarization reversal in FE layer is a promising way to induce a threshold switching behavior and has been successfully used to mimic neuronal firing activities. In nanoscale Si-doped HfO2-based FeFETs, two key neuronal dynamics have been demonstrated. Integration of action potentials and the subsequent firing were optimized in these devices by designing specific electrical excitations that induce particular NLS kinetics of the FE layer. To tune the firing frequency, arbitrary refractory period, and the leaky effect, \(E_{dep}\) needs modulation as well because it dictates the time during which the FE polarization will start to relax, i.e., the leak effect will take place. The FeFET-based LIF neuronal behavior was reported from 30 nm (channel length) \(\times 80\) nm (width) devices\cite{30} that highlights the possibility of advanced down-scaling of neuron circuits which can significantly reduce the complexity of neuromorphic architectures.

6. Application in Neuromorphic Architectures

6.1. Ferroelectric Memories in Reprogrammable Arrays

A significant part of the DNN processing is VMM operation between the input vector (voltage pulses) and the weight matrix (conductance of the memristors), which essentially performs multiply-and-accumulate (MAC) operation and generates an output current vector. In this respect, IMC is a promising paradigm as it performs the computation directly into memory subarrays and in-principle can be performed in fully parallel fashion. In a recent work by Toshiba scientists\cite{37} it was shown that nonlinear FTJ memristors can perform linear computation at ultralow currents (Figure 9). Using Si-doped HfO2-based FTJ crossbars, an analog-voltage-amplitude VMM was demonstrated. By utilizing the device nonlinearity, that remains constant for multiple conductive states, a selector-less array is obtained. This result opens possibility to exploit these features for analog VMM-intensive applications, such as neural inference engines, toward energy efficiencies above 100 TOPS W\(^{-1}\).
6.2. FeFET-Based BCNNs

Chen et al.\textsuperscript{[138]} reported a FeFET crossbar array-based binary convolutional neural network (BCNN) where the basic cell consisted of two FeFETs and two access transistors. In this architecture, two FeFETs, storing complementary bits, represent one weight bit. FeFET crossbar designs of array size 64 \times 64, simulated using the HSPICE FeFET models, showed write power reductions of 5600\times and 395\times compared to RRAM and CMOS equivalents, depending on the array design. Read power can be reduced by 4.1\times and 3.1\times. However, the read latency is found to be 8\% higher with the FeFET approach. In overall comparison with a CMOS-based design, the FeFET-based design showed best performance in read power-delay product (PDP), indicating a promising future direction to explore and benchmark.

Figure 9. a) Schematics of the FTJ device structure and transmission electron microscopy (TEM) image showing the material stack. b) Parallel line-by-line programming strategy for a 5 \times 5 selector-less (passive) FTJ crossbar reminiscent of the V/2 scheme. \( V_p \) is the amplitude of the wordline biphasic write pulse, and is lower than the \( V_{th} \) of the FTJ. c) Demonstration of the pulse-induced analogue bipolar switching of the FTJs. d) Comparison of a read operation between the FTJ devices and other memory devices utilized in linear VMM operations showing potential ultralow energy consumption. \( E_R \) and \( E_C \) are energy consumptions by the resistive and capacitive parts, respectively. Device indicators correspond to: ReRAM1, ReRAM2, and PCM. e) Illustration of a multilayer neural network trained for the F-MNIST task to classify 28 \times 28 pixel images of garment items into ten different categories. A neural network in hardware implemented by FTJ crossbars for the weight layers can operate fully analog apart from the peripheral digital converters. X and Y represent the input value vector and output label vector, respectively. f) Simulations of inaccurate weight transfer of the network trained on F-MNIST to FTJ crossbars implementing synaptic weight layers. Demonstrated programming precision for \( >8 \) states with \( \sigma = 3.5\% \) would reduce the network’s accuracy by 1\%. Thirty-two separate states programmed within \( \sigma = 2\% \) would recover the initial test accuracy. g) Computation efficiency estimations via simulations of the network. Reproduced with permission.\textsuperscript{[37]} Copyright 2020, Springer Nature.
6.3. Application in DNN and SNN

DNNs are one of the most used ANNs where training and inferencing are performed based on backpropagation algorithm. Working principles of DNNs are not quite biologically inspired, as, the neuron outputs in a DNN are still real-valued numbers, processed on synchronous time-steps. However, massively parallel VMM operation is at the heart of DNN forward-inferencing and training. In a DNN, integration of device read currents along the columns of a crossbar array implements forward propagation of neuron excitations while integration along rows implements the backpropagation of error terms. From material and device perspective, ideal characteristics for more than 90% DNN training accuracy require linear and symmetric conductance change (both potentiation and depression wise) of the synapses with identical stimulating pulses, large dynamic range of conductance, and minimum variability of conductance states. Some degree of nonlinearity is tolerable, if the conductance range over which the response is nonlinear is only a small fraction of the overall conductance range. Jerry et al.[129] showed with FeFET-based 5-bit analog synapse with symmetric potentiation and depression characteristics, it is possible to attain \( \frac{G_{\text{max}}}{G_{\text{min}}} \) \( \approx 90\% \) DNN online training accuracy. However, instead of an identical pulse scheme, an incremental pulse scheme is used to achieve this accuracy. Kim et al.[67] studied the influence of the FE properties and contact barrier heights in the FeFET synapses based on P(VDF-TrFE) on the training and recognition performance of a multilayer perceptron (MLP)-based neural network (Figure 10). Two different datasets from the MNIST database, i.e., the fashion and handwritten digits, were used for the training and inference tasks. Investigation on the effect of annealing temperature of P(VDF-TrFE) and the nature of the contact metals (Ti, Cr, Pd) of the FeFET showed excellent FE properties with maximized amount of \( \beta \)-phase and a metal, that forms a relatively high tunneling injection barrier, can improve the dynamic conductance range (DR), defined as ratio of maximum and minimum conductance \( \left( \frac{G_{\text{max}}}{G_{\text{min}}} \right) \) and nonlinearity of the postsynaptic current (PSC). This improvement in dynamic range and nonlinearity by suppression of tunneling current can result in improved accuracy of recognition in neural networks. The effect of the contact metal on the recognition rate revealed that the synapses with Pd contact can achieve maximum recognition rate of 75.2% which is about 10% lower in comparison to that of software NNs (86.8%), whereas the rates obtained with Ti- or Cr-contacted synapses were less than 50%. This study opens the prospect of optimizations of P(VDF-TrFE) morphology, semiconductor–FE interface, and contact metals for high-performance cognitive tasks based on P(VDF-TrFE) FeFETs.

In SNNs, neurons operate on asynchronous, uni-valued, event-based spikes while learning takes place in the synapses by STDP-based local Hebbian learning rule. SNNs are efficient in learning spatiotemporal pattern in data and therefore from FE

![Image](https://www.advancedsciencenews.com/)

**Figure 10.** a) Schematic illustrations of MNIST fashion dataset and b) constructed multilayer perceptron neural network. c) Recognition rate as a function of the number of epochs with respect to annealing temperatures for P(VDF-TrFE) film. d) LTP postsynaptic current (PSC) responses as a function of the number of spikes for different dynamic range (DR) (upper panel). Recognition rate as function of the corresponding DR value (lower panel). e) Recognition rate as a function of the number of epochs with respect to contact metals. f) Recognition rate as functions of DR (upper panel) and LTP nonlinearity (lower panel). Reproduced with permission.[67] Copyright 2021, Royal Society of Chemistry.
device perspective, one important aspect is properly engineered FE polarization and depolarization dynamics to mimic the LIF neuron firing and implementing accumulative nature of switching for STDP-based synaptic weight update. Similar to DNNs, device-level randomness and limited bit precision of on-chip synaptic weights also affect the classification accuracy of SNNs adversely. In all FeFET-based SNNs with HZO synaptic and neuron elements, Dutta et al. showed a stochastically firing neuron helps improve the SNN classification accuracy in the presence of reduced weight precision. In general, reduced bit precision results in reduced accuracy, i.e., up to 6-bit precision, a test accuracy of 95.4% is seen that starts decreasing to 91% for 5 bits and drastically down to 43.5% for 4-bit precision. However, upon adding noise, the accuracy for 5–8 bits increased to 96%. For 4-bit weights, the accuracy improved substantially with more noise around the threshold; however, starting from 3-bit precision, no improvement was found, emphasizing the importance of higher bit precision for accurate learning in SNNs. Simulations based on both HfO2-based FTJ and junctionless (JL) FE FinFET showed pattern recognition accuracy to be ≈80%. This moderate accuracy rate, in both cases, is a consequence of nonideal conductance update under identical pulse schemes. An incremental pulse scheme, coming at the cost of circuit complexity, showed improvement in recognition accuracy to >90% level. In a recent work, it was shown that shorter pulses of 20 ns width is able to improve weight update linearity in FTJ synapses under identical pulse scheme. Also, addition of an ultrathin dielectric capping layer on the FE layer in a FeFET can make the polarization rotation more gradual due to linearly regulated gate stack capacitance and has the potential to improve bit precision and weight update linearity. Such custom-designed pulse schemes or innovative gate stack design, with proper combination of thickness, dielectric constants ($\varepsilon_{\text{dielectric}}$) and gate geometries, improvement in linearity, and bit precision, can be obtained improving both recognition accuracies in SNNs or online training in DNNs.

### 6.4. Ferroelectric Devices for Wearable Intelligent Systems

Synaptic functionality under extreme bending conditions was demonstrated in ultrathin conformable organic synaptic transistors based on P(VDF-TrFE) and pentacene (Figure 11). The

![Figure 11](https://www.advancedsciencenews.com)

**Figure 11.** Synaptic characteristics of the flexible pentacene–PVDF-based FeFETs on different uneven surfaces and under extreme bending conditions. a) Schematics of device architecture, PVDF, and pentacene chemical structure and images of the conformable FeFETs on different uneven surfaces, device on/off ratio and MW on different surfaces. b) Images of devices on brain-shaped PDMS mold and FeFETs folded with a bending radius of 50 μm and c) LTP and LTD behavior of the FeFETs on the brain mold (red line) and the in folded condition (black line). d) PSC response of the folded devices was found to depend on the pulse width and the number of pulses. e) Repetitive LTP and LTD operations in the folded condition during 6000 spikes of presynaptic pulses. One cycle of 30 potentiating pulses was followed by subsequent 30 depressing pulses. Reproduced with permission. Copyright 2019, American Chemical Society.
device on/off ratio, MW, and synaptic characteristics of the fabricated free-standing devices on various uneven substrates, such as thermal-shrink plastic film, jelly, textile, candy, toothbrush, and a brain-shaped PDMS mold, were tested. Also, the FeFETs were tested under extreme bending conditions, i.e., up to a bending radius of 50 μm. The fabricated synaptic transistors exhibited stable LTP/LTD features under both normal and harsh bending conditions. This performance stability suggests a new research direction for realization of wearable artificial intelligent systems. Lee et al.\cite{Lee2020} reported a flexible, artificial, tactile sensory organ with intrinsic–synaptic functionalities. For this application, a synaptic transistor with P(VDF-TrFE):BaTiO3 nanocomposite as the gate dielectric and pentacene as semiconducting channel was used. Triboelectric-capacitive coupling under finger touch induced FE dipole switching that modulates the postsynaptic current of the device and allowed reception and slow adaptation. Emulation of synapse-like connection of receptor cells and afferent neuron terminals in a sensory organ enables implementation of the sensor with intrinsic intelligence without link to a neuronal processor and thus defines a new paradigm in sensor technology for neurorobotics, autonomous systems, intelligent electronic skins, and prosthetics (Figure 12).

6.5. Circuit Integration Issues

Write disturbance or nearest neighbor cross-talk is one of the key issues for the dense integration of memories. Write disturbance

![Figure 12. Intrinsically intelligent artificial tactile sensory organ with sensing and processing ability. The organ comprises adapting sensing mechanism and a synapse-like connection with an afferent neuron. Pressure induces influx of Ca^{2+} ion through the piezo channel increasing the membrane potential. The increase in membrane potential eventually leads to synaptic firing through release of neurotransmitters. Synaptic firing occurs in response to tactile information and is processed by synaptic plasticity. a) Schematic representation of dipole switching in the FE gate layer and accumulation of holes in the FeFET channel induced by tactile stimulation (I–IV). Illustration of formation principle of P_r with a change in channel conductance corresponding to synaptic weight. Relative changes in the PSC of the FeFET at b) different touch pressures, c) duration (≈1 kPa), and d) touch rate (≈1 kPa for 5 s). e) Reception by a flexible sensory organ occurs in a self-energy transducing way by triboelectric effect during finger touch. This eventually modulates the drain current with aligned dipoles in the BaTiO3 NPs/P(VDF-TrFE) nanocomposite by generated triboelectric capacitive potential, resulting in a sensing and processing at the same node. Reproduced with permission.\cite{Lee2020} Copyright 2020, The Author(s). Published by Springer Nature, Creative Commons CC BY license.](https://www.advancedsciencenews.com)
appears when a programming voltage across a certain column or row not only affects the selected memory cell but also affects the neighboring columns or rows causing a memory cell to deviate from its programmed state. Even though mitigation strategies are available with extra components like selectors, the ideal goal is always to minimize this cross-talk within single memory cell. FTJs with one semiconducting electrode or a FE diode can provide a large nonlinearity factor in a device $I–V$, leading to a selectivity in the direction of current flow, lowering the risk of cross-talk. Another important aspect of FE systems is the large voltage-timescale dependence. In P(VDF-TrFE)-based FTJs, a seven-order-of-magnitude change in switching timescale is reported at the expense of doubling the programming voltage. This has important implications in terms of integration of FTJs in passive crossbar arrays. When the programming timescale of a memory cell varies by orders of magnitude due to a slightly less programming voltage, the array can become intrinsically robust against write disturbance. For instance, a device switching at 6 V with a 20 ns programming pulse will not switch a neighboring cell with an inadvertent 4 V pulse at the same timescale. Therefore, through material and device design, a selector-less integration of FE memories in large crossbar arrays could be possible, greatly improving the memory integration density, without affecting the reliability of the programmed memory matrix. For doped hafnia-based FTJs, a low readout current in ultra-scaled devices is an issue of major concern. Current density of FTJs, with 10 nm HZO, is $\approx 10^{-4}$ A cm$^{-2}$. A crossbar array using this current density for the 20 nm technology node, can give an on-state current $\approx 0.7$ nA. In a 1024 $\times$ 1024 array with all the WLS turned on and all the devices in the on state, the BL summed current will still be below 1 pA. Therefore, in a 1024 $\times$ 1024 array, considering $R_{\text{wire}}$ and $C_{\text{wire}} = 2.925 \Omega$ and 8.8 aF, respectively, the first-order estimation of RC is larger than 1 s that can make the readout extremely slow. To overcome this issue, current values from an ultrathin 1 nm HZO is considered. The projected on-state current for 1 nm HZO is equivalent to about 60 pA for a device at 20 nm node. Innovative design concept of stacked capacitors from modern DRAM processes, with a diameter to height aspect ratio of 220, is proposed as a viable choice. Following this idea, a crossbar array in the 20 nm node could show a desirable summed current level on the order of 10 μA. To model a realistic data pattern of on-state and off-state devices in a 1024 $\times$ 1024 array, quantized weights from a fully connected layer of a DNN are mapped on the FTJ cells. Evaluation of summed current, accuracy, and delay suggests that low current issues in an FTJ crossbar array can be overcome utilizing innovative design concepts.

7. Integrated Ferroelectric Memories for Edge Applications

Miniaturized sensor systems with built-in memory and computing functionalities are becoming the cornerstones of AI at the edge. Construction of such hardware would require functional devices in which synaptic information processing capability will be merged with various sensing functionalities such as sensitivity to light, pressure, temperature, chemical analytes, etc. The most efficient way to do this is a system-level integration of sensing and computing in electronics. Recent years have witnessed a surge in research efforts for such platform where either a 2D, 2.5D or 3D integration of layers with different functionalities has been demonstrated.

7.1. CMOS Integrated HfO$_2$-Based FRAM, FTJs, and FeFETs

In recent times, monolithic 3D integration of HZO-based capacitors, FTJs, and FeFETs with FEOL high-k/metal gate Si-NOMOSs has been shown (Figure 13). In all cases, a low thermal budget (maximum 450 °C, 80 s) processing is used to integrate HZO devices on CMOS platform. Francois et al. demonstrated scalability of TiN/HZO/TiN capacitors from 100 μm diameter down to 300 nm by coincorporating the capacitors in the BEOL of 130 nm CMOS technology. Excellent performance, like $2P_R > 40 \mu$C cm$^{-2}$, endurance $>10^{13}$ cycles, switching speeds $<100$ ns, operating voltages $<4$ V, and data retention at 125 °C, is reported on the scaled bit cells. Apart from the single devices, 16 kbit 1T–1C arrays with capacitors of different sizes were demonstrated where programming and erasing were performed by voltage pulses and readout was done with an integrated sense amplifier.

For HZO FeFETs, Dutta et al. used a 1% Tungsten (W)-doped amorphous In$_2$O$_3$ semiconducting channel material. A high remnant polarization charge density $2P_R$ of 40 μC cm$^{-2}$ with reliable switching characteristics was demonstrated in ultrascaled FeFETs with 20 nm channel length. A read MW of 0.45 V, write speed of 100 ns, and write endurance $>10^8$ cycle, together with 2-bit per cell synaptic weight with well-separated conductance states, were demonstrated. System-level analysis of IMC accelerators for performing inference on CIFAR-10 image dataset predicts that 22 nm BEOL FeFET can achieve $3 \times$ higher energy efficiency compared to the 7 nm SRAM while occupying a smaller area due to area folding enabled by 3D integration.

Wafer-scale uniformity of device performance is an essential criterion for implementation in large-scale circuits. Study has been made on large-area device performance for both hafnia- and P(VDF-TrFE)-based FeFETs. Jiang et al. showed a very tight distribution of device variability across 300 mm wafer of HZO capacitor, while data from Global Foundries confirm similar performance from the $I–V$ curves of FeFET passive arrays from across wafer programmed into 63-bit $L = 450$ nm, W = 450 nm co-integrated in 28 nm technology (Figure 13e). Device yield was reported to be $>90\%$. Results show that the scan-to-scan variability increases with decreasing junction size due to inherent variability in FE domain switching. For P(VDF-TrFE)-based FeFETs, large-area printed memory transistors with highly uniform device performance were reported in the literature. Bae et al. demonstrated roll-up memory arrays where Sekitani et al. developed large-area flexible communication sheet with ultralow power consumption. van Breemen et al. reported 16 × 16 passive matrix of P(VDF-TrFE)-based FeFET arrays where current modulations of $\approx 10^5$ retention time of more than 12 days and fault tolerant writing were demonstrated. In all cases, printing technique was used for the FE memory array fabrication. With innovative nanoimprint lithography technique, a memory cell density of 33 GB in $2^2$ was reported from P(VDF-TrFE) memories by Hu et al., which opens pathways for dense integration of these devices.
7.2. Ferroelectric Polymer-Based Memory Integrated Sensors for Smart and Adaptive Systems

Integration of flexible sensing and memory arrays has been demonstrated in the past, mainly for touch or light sensing and remembering. For instance, Sekitani et al. first demonstrated the integration of flexible pressure sensors and memory arrays on PEN substrates.[151] Human visual perception systems were emulated by integrating memristors with UV image sensors on flexible PI substrates.[28] In the artificial visual memory system, the detection of UV light by the image sensors was recorded in memristors. This provided detection and memory capacities that provided emulation of human visual memory. In another approach, a FE/electrochemical modulated organic synaptic device was integrated with organic light-sensitive elements showing improved NVM functions and three types of synaptic plasticity.[152] Organic neuromorphic devices for artificial neural network applications have been discussed in detail in some recent works.[23,53,153] Ferroelectric polymers are promising for vital parameter sensing like pulse pressure and change in body temperature.[154] By combining their piezo- and pyroelectric properties, multiple parameter sensing through a single sensing element is possible.[155,156] Additionally, the nonvolatile polarization switching in these devices offers the possibility of integrating the sensing and memory performance in the same device architecture. Earlier results showed an in situ poled P(VDF-TrFE) nanowire on an Au-coated PI film enabled detection of finger motion, breathing, and pulse rate in humans. Ferroelectric skins with pressure sensing layer made of two interlocked microdome sheets for monitoring artery pulse pressure waveforms at different skin temperatures (20–40 °C) before/after exercise showed the radial artery augmentation index and the radial diastolic augmentation index vary linearly with temperature.[157] These results from a promising baseline and suggest future directions for improving sensing accuracy for integration with IMC architectures. Additionally, an FeFET with photo, gas, or humidity sensing semiconductor can further broaden the horizon of sensing, while FE gate stack brings the memory and computing possibilities in the same device, making in-sensor computing possible.

8. Future Perspective and Outlook

Despite the fast recent progress of the low-temperature processable, FE-based NVM technology, several issues remain to be solved before their true potential is fully realized. One of the main being the low-voltage operation while keeping their best performance intact. Hafnia-based FeFET technology has already
reached a considerable matured stage of research and therefore can be predicted to become one major contributor in the field of CMOS compatible NVM and neuromorphic computing circuits. A few recent results showed promising way forward on solving the issues like programing voltage MW bottleneck and device breakdown after a million operating cycles. The field of FTJs is, however, still in its genesis, and there is high potential for improvement both from technological and fundamental perspectives. For scaled devices, randomness in switching voltages is still an issue. A targeted programming scheme has been predicted to show improved performance in this respect.\[^{[146]}\] Another point of concern is precision of analog conducting states that can be detrimental for training accuracy. Future research in the direction of innovative gate stack and geometry design or multilayer tunnel barrier design can be useful. However, many brain-like computing architectures rely on statistical approximations and estimations of data and therefore error-free bit precision might not always be essential. Also, device aware architecture and algorithm designs can alleviate some issues to produce high throughput, noise immune, high density CIM arrays.\[^{[158]}\] For organic FE devices, a high-voltage operation in many earlier reports was considered to be a hurdle for their utilization in wearable technologies. However, some promising but largely overlooked results suggest that through nanopatterning,\[^{[150,159]}\] gate stack thinning and optimization,\[^{[160]}\] and use of high mobility semiconductor components, the operating voltages can be brought down drastically. As there are no fundamental limitations for their fatigue or breakdown under continuous bias stress, it is possible to achieve more than 10\(^5\) cycles of operations in these devices, making them potentially suitable for online training in IMC circuits. It is also possible to take care of many device-level nonidealities at the circuit level.\[^{[160]}\] Need for ultrastretchable and bendable devices for wearable, prosthetics and artificial skin applications can be met with the flexible polymeric devices opening a new path for low-cost, large-area intelligent device applications. Ubiquitous sensing and computing platform utilizing the FE materials would be the next target to avoid the need for power-and-area hungry analog-to-digital or reverse signal conversion in the edge computing devices and lead the way toward realization of a truly bioinspired intelligent systems.

Acknowledgements

The author acknowledges financial support from European Union’s Horizon 2020 research and innovation program under grant agreement no. 101016734 and discussions with Arto Rantala and other ASIC designer colleagues at VTT.

Conflict of Interest

The author declare no conflict of interest.

Keywords

adaptive sensing, edge computing, electronic synapses, ferroelectric field-effect transistors, ferroelectric tunnel junctions, neuromorphic computing, neurons

Received: June 16, 2021
Revised: October 29, 2021
Published online: December 8, 2021

[1] Beyond von Neumann, Nat. Nanotechnol. 2020, 15, 507.
[2] Nat. Electron. 2020, 3, 347, https://doi.org/10.1038/s41928-020-0457-1.
[3] W. Zhang, B. Gao, J Tang, P. Yao, S. Yu, M.-F. Chang, H.-J. Yoo, H. Qian, H. Wu, Nat. Electron. 2020, 3, 371.
[4] O. Krestinskaya, A. P. James, L. O. Chua, IEEE Trans. Neu. Net. Learn. Syst. 2020, 31, 4.
[5] A. Mehonic, A. Sebastian, B. Rajendran, O. Simeone, E. Vasilaki, A. J. Kenyon, Adv. Intell. Syst. 2020, 2, 2000085.
[6] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, E. Eleftheriou, Nat. Nanotechnol. 2020, 15, 529.
[7] Q. Xia, J. J. Yang, Nat. Mater. 2019, 18, 309.
[8] Y. Li, K.-W. Ang, Adv. Intell. Syst. 2021, 3, 2000137.
[9] F. Alibart, E. Zamanidoost, D. B. Strukov, Nat. Commun. 2013, 4, 2072.
[10] M. Prezioso, F. Merrick-Bayat, B. D. Hoskins, G. C. Adam, K. K. Lakhe, D. B. Strukov, Nature 2015, 521, 61.
[11] C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, Nat. Commun. 2018, 9, 2385.
[12] G.W. Burr, R. M. Shelby, S. Sidler, C. di Nolofo, J. Jang, I. Boybat, R.S. Shenoy, P. Narayanan, K. Virwani, E.U. Giacometti, B.N. Kurdi, H. Hwang, IEEE Trans. Elect. Dev. 2015, 62, 3498.
[13] S. Yu, Z. Li, P.-Y. Chen, H. Wu, B. Gao, D. Wang, W. Wu, H. Qian, in IEEE International Electron Devices Meeting (IEDM), IEEE, Piscataway, NJ 2016, p. 16.2.1.
[14] C. Lee, P. Panda, C. Srinivasan, K. Roy, Front. Neurosci. 2018, 12, 435.
[15] S. F. Cooke, T. V. P. Bliss, Brain 2006, 129, 1659.
[16] I. F. Abbott, S. B. Nelson, Nat. Neurosci. 2000, 3, 1178.
[17] D. Ielmini, S. Ambrogio, Nanotech. 2020, 31, 092001.
[18] M.-K. Kim, J.-S. Lee, Nano Lett. 2019, 19, 2044.
[19] S. Dai, Y. Zhao, Y. Wang, J. Zhang, L. Fang, S. Jin, Y. Shao, J. Huang, Adv. Funct. Mater. 2019, 29, 1903700.
[20] S. Oh, H. Hwang, I. K. Yoo, APL Mater. 2019, 7, 091109.
[21] H. Li, R. Wang, S.-T. Han, Y. Zhou, Polym. Int. 2020, 69, 533.
[22] M.-K. Kim, Y. Park, I.-J. Kim, J.-S. Lee, iScience 2020, 23, 101846.
[23] Y. van de Burgt, A. Melianas, S. T. Keene, G. Malliaras, A. Salleo, Nat. Electron. 2018, 1, 1386.
[24] H.-M. Huang, Z. Wang, T. Wang, Y. Xiao, X. Guo, Adv. Intell. Syst. 2020, 2, 2000149.
[25] M. D. Pickett, C. Medeiros-Ribeiro, R. S. Williams, Nat. Mater. 2013, 12, 114.
[26] P. Stoliar, J. Tranchant, B. Corrza, E. Janod, M.-P. Besland, F. Tesler, M. Rozenberg, L. Cario, Adv. Funct. Mater. 2017, 27, 1604740.
[27] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, E. Eleftheriou, Nat. Nanotechnol. 2016, 11, 693.
[28] Z. Wang, S. Joshi, S. Savelev, W. Song, R. Midya, Y. Li, M. Rao, P. Yan, S. Asapu, Y. Zhuo, H. Jiang, P. Lin, C. Li, J. H. Yoon, N. K. Upadhyay, J. Zhang, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, R. S. Williams, Q. Xia, J. J. Yang, Nat. Electron. 2018, 1, 1387.
[29] S. Majumdar, H. Tan, I. Pande, S. van Dijken, APL Mater. 2019, 7, 091114.
[30] H. Mulaosmanovic, E. Chicca, M. Bertele, T. Mikolajick, S. Slesazeck, Nanoscale 2018, 10, 21755.
[31] F. Cai, J. M. Correll, S. H. Lee, Y. Lim, V. Bothra, Z. Zhang, M. P. Flynn, D. W. Lu, Nat. Electron. 2019, 2, 290.
[32] W.-H. Chen, C. Dou, K.-X. Li, W.-Y. Lin, P.-Y. Li, J.-H. Huang, J.-H. Wang, W.-C. Wei, C.-X. Xue, Y.-C. Chiu, Y.-C. King, C.-J. Lin,
Sayani Majumdar is a senior scientist in microelectronics and quantum technologies at the VTT Technical Research Centre of Finland and adjunct professor at University of Turku. She received her Ph.D. in 2006 from Jadavpur University and worked as a postdoctoral fellow at the Åbo Akademi University, University of Turku and in Aalto University as Academy of Finland research fellow. She was a visiting scientist at Massachusetts Institute of Technology and Max-Planck Institute for Polymer Research. Her current research focuses on CMOS and flexible substrate compatible heterostructures for neuromorphic computing and adaptive sensing.