Letters

1F-1T Array: Current Limiting Transistor Cascoded FeFET Memory Array for Variation Tolerant Vector-Matrix Multiplication Operation

Masud Rana Sk, Sunanda Thunder, Franz Müller, Nellie Laleni, Yannick Raffel, Maximilian Lederer, Luca Pirro, Talha Chohan, Jing-Hua Hsuen, Tian-Li Wu, Konrad Seidel, Thomas Kämpfe, Sourav De, and Bhaswar Chakrabarti

***Abstract*—This letter proposes a memory cell, denoted by 1F-1 T, consisting of a ferroelectric field-effect transistor (FeFET) cascaded with another current-limiting transistor (T). The transistor reduces the impact of drain current (I_d) variations by limiting the on-state current in FeFET. The experimental data from our 28 nm high-k-metal-gate (HKMG) based FeFET calibrates and simulates the memory arrays. The simulation indicates a significant improvement in bit-line (BL) current (I_{BL}) variation and the accuracy of vector-matrix multiplication of the 1F-1 T memory array. The system-level in-memory computing simulation with 1F-1T synapses shows an inference accuracy of 97.6% for the MNIST hand-written digits with multi-layer perceptron (MLP) neural networks.**

***Index Terms*—1F-1T, FeFET, HfO_2, memory array, vector-matrix-multiplication.**

I. Introduction

Deep neural networks (DNN) play a significant role in performing many data-intensive computing tasks such as speech recognition, motion detection, computer vision, and natural language processing. Training DNNs with enormous amounts of data from the internet and real-time devices leads to high energy and latency costs. Recently, in-memory-computing (IMC) with emerging non-volatile memory (eNVM) technologies are being researched to alleviate this issue [1], [2], [3], [4], [5], [6]. The primary step of IMC is vector-matrix multiplication (VMM). Many eNVMs such as resistive random access memory (ReRAM) [7], [8], [9], [10], [11], [12], [13], phase change memory (PCM) [14], and FeFETs [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27] have been investigated in recent years.

Amidst such a plethora of eNVMs, FeFET is a promising one due to field-based operation, low power consumption, fast switching, high on-current (I_{ON}) to off-current (I_{OFF}) ratio ($\frac{I_{ON}}{I_{OFF}}$), excellent linearity and bidirectional programmability, good endurance, and compatibility with the Complementary Metal-Oxide Semiconductor (CMOS) technology [18], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38]. However, the primary obstacle in implementing FeFET-based computing systems lies in the inherent stochasticity of FeFET devices. One of the primary reasons for this stochasticity lies in the polycrystalline nature of HfO_2 -based ferroelectric thin films. The presence of charge traps at the Ferroelectric /interlayer interface and within the Ferroelectric film can lead to asymmetrical conductive response and large device-to-device variations [39], [40], [41]. Great efforts have been made to minimize the effects of such non-idealities both from the devices and from the perspective of the circuit [42], [43], [44], [45], [46], [47]. It is imperative that device-to-device variations in the drain currents of a FeFET can adversely affect the performance of a synaptic core, resulting in significant degradations in training and inference accuracy. Our previous work demonstrated how a series-resistor connected to the drain terminal of the FeFETs could reduce variations by limiting the current in the Low-Voltage-Threshold (LVT) state [21]. However, the fabrication of resistors in a standard CMOS process adds complexity to the macro-design because of the enormous size of such resistors. Poly-silicon resistors have the highest resistance density but suffer from a larger mismatch, requiring 100 μm^2 to achieve 1Mohm.

In this work, we demonstrate the efficacy of the 1F-1T structure in overcoming this issue. We have considered 4 × 4 arrays to evaluate the effects of this 1F-1T architecture. We show that such a configuration prevents the accumulation of variations over bit-line current (I_{BL}) and also reduces the impact of voltage swing across word lines (WL), bit lines (BL), and select lines (SL). We benchmark the system-level performance of an in-memory computing circuit with a 1F-1T synaptic array. Our simulations indicate significant improvements in inference accuracy compared to the network using 1F synapses.

Manuscript received 1 June 2023; accepted 3 July 2023. Date of publication 14 July 2023; date of current version 28 July 2023. This work of Masud Rana Sk and Bhaswar Chakrabarti was supported in part by the Research Center on Advanced Memory and Computing under the IoE scheme at the IIT Madras, and in part by the European Union's ECSEL under Grants 826655 -Project TEMPO and 876925 Project ANDANTE. The review of this letter was arranged by Associate Editor J. -J. Huang. *(Masud Rana Sk and Sunanda Thunder contributed equally to this work.) (Corresponding authors: Sourav De; Bhaswar Chakrabarti; Tian Li Wu.)*

Masud Rana Sk and Bhaswar Chakrabarti are with the Department of Electrical Engineering, IIT Madras, Chennai 600036, India (e-mail: bchakrabarti@ee.iitm.ac.in).

Sunanda Thunder, Franz Müller, Nellie Laleni, Yannick Raffel, Maximilian Lederer, Konrad Seidel, Thomas Kämpfe, and Sourav De are with Fraunhofer-IPMS, CNT, 01109 Dresden, Germany (e-mail: sourav.de@ipms.fraunhofer.de).

Luca Pirro and Talha Chohan are with GlobalFoundries, 01109 Dresden, Germany.

Jing-Hua Hsuen is with IPSI, NYCU, Hsinchu 30010, Taiwan.

Tian-Li Wu is with IPSI, NYCU, Hsinchu 30010, Taiwan, and also with ICST, NYCU, Hsinchu 30010, Taiwan (e-mail: tlwu@nycu.edu.tw).

Digital Object Identifier 10.1109/TNANO.2023.3295093

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

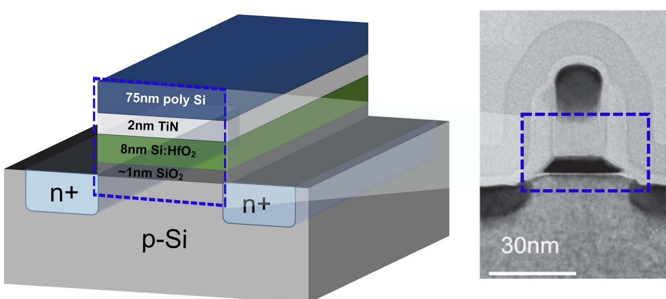

Fig. 1. Schematic and corresponding TEM image of a 28 nm HKMG-based FeFET fabricated at GolbalFoundries. [26].

II. Experiments

A. Characterization of 28 nm HKMG FeFET

The experiment began with fabricating the crossbar arrays and memory cell structures on 300 mm wafers using the 28 nm HKMG technology at GlobalFoundries. Fig. 1 shows the schematic of an HKMG FeFET and associated transmission electron microscopic (TEM) image of a minimum feature size FeFET at this technology node. On the application of voltage pulses to the gate terminal, the ferroelectric layer's dipoles align themselves as per the polarity of the pulse, altering the channel's surface charge density and, consequently, the threshold voltage (V_{th}). A positive pulse at the gate terminal of n-type FeFET programs the device at a lower V_{th} state (LVT), and the negative pulse programs the device at a higher V_{th} state (HVT). The devices were programmed and erased by applying 500 ns pulses of amplitude 4.5 V and -5 V, respectively, to the gate terminal of FeFET, with the drain and source terminals grounded. Before READ-WRITE operations, the devices were subjected to wake-up cycling (100 times), delivering 4.5 V and -5 V pulses of 500 ns.

B. Design of 28 nm HKMG Based Memory Array

In this work, the effects of FeFET variability have been assessed using 4×4 memory arrays. We investigated the two distinct architectures shown in Fig. 2. Note that program and erase conditions for the single devices were adopted to simulate the characteristics of the array. Fig. 2(a) shows the architecture with one FeFET device per synaptic cell (1F architecture). Here, word lines (WL) are connected to the gate terminals of the FeFETs in a row. Each cell can be accessed by controlling the corresponding WL and the select line (SL). Bit-line (BL) current (I_{BL}) defines the state of the cell. Our proposed schematic of the 1F-1T synaptic array is shown in Fig. 2(b). Here, each synaptic cell consists of an extra transistor which acts as a current limiter. A bias voltage is applied to the transistor's gate terminal to access a specific cell. FeFET conductance is controlled by tuning the WL and SL voltages. Layouts of the 4×4 1F and the proposed 1F-1T memory array are shown in Fig. 3. The 1F and 1F-1T memory array have an area of 5580 λ^2 and 7688 λ^2, respectively. Compared to the pseudo crossbar 1F- 1T memory array [16], our proposed 1F-1T memory provides a 60% area advantage.

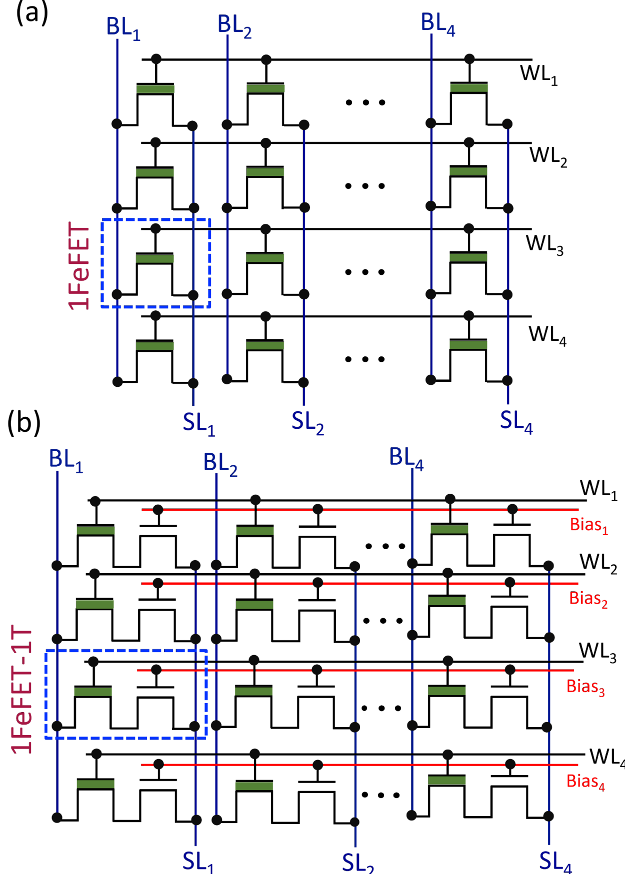

Fig. 2. 4×4 memory arrays built with (a) 1F based synaptic cells and (b) with 1F-1T synaptic cells.

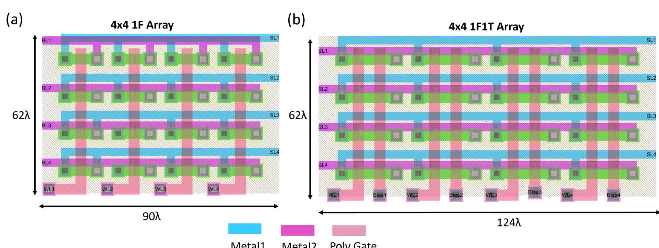

Fig. 3. Layout of (a) 1F memory array and (b) proposed 1F-1T memory array corresponding to the schematic shown in Fig. 2.

TABLE I
Pulse Amplitude in Synaptic Array

	WL	BL	SL	Bias
Write	4.5 V	0 V	0 V	1.8 V
Read	0-1.5 V	0.1 V	0 V	0.3 V
WL/Bias Inhibit	-0.3 V	x	x	0 V

This is due to the same diffusion layer shared by the cascaded transistor's drain and source of FeFET in our configuration.

Table I summarizes the operating voltages for simulating the 1F and 1F-1T arrays. The row-wise write operation was conducted by applying 4.5 V at the word line (WL) and 1.8 V at

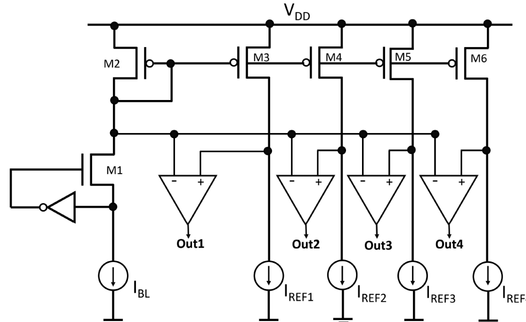

Fig. 4. Block scheme of Current Sense Amplifier (CSA) for sensing the four stages bit line current (I_{BL}).

the Bias line. The cell-wise read operations were conducted by biasing BL to 0.1 V, WL at 0-1.5 V, and Bias line at 0.3 V. While writing/reading operations are performed on a particular row, all other rows are inhibited by applying -0.3 V to WLs and 0 V to the Bias lines. Similarly, columns are disabled by controlling the SL voltages so that the total gate-to-source (V_{GS}) is below the threshold voltages of the FeFETs. The operating voltages are given in Table I for the write, read, and inhibition operations in the arrays.

The current sense amplifier (CSA) reads the different levels of bit-line current. Fig. 4 depicts a typical 4-stage current-mode sensing amplifier [48], [49]. I_{BL} and I_{REF} stand for the memory array's bit-line current and reference current, respectively. 'Out' denotes the differential comparator's logical output voltage. M2 mirrors the bit-line current into the reference branches, which are the drain sides of M3, M4, M5, and M6. Current-to-voltage conversion takes place at each reference node for the reference current. The output node Out rises to logic high 'V_{DD}' if the memory array's I_{BL} exceeds the I_{REF}. Otherwise, it remains at logic low '0'.

C. Neural Network Simulation

Finally, the impact of variations on the system-level performance, especially for in-memory computing applications, was evaluated by the Neurosim platform [50]. Experimentally calibrated I_d values with the statistics of variations were used to emulate the synaptic weights of a multilayer perceptron (MLP) neural network (NN). MNIST dataset is used to benchmark the performance of the MLP. The neural network's architecture is illustrated in Fig. 5. This work considers offline training for neural networks, where synaptic weights are pre-trained in the software. The weights are then encoded into the circuitry. This method is more energy-efficient but less noise-tolerant as the synaptic weights cannot be modified on-the-fly during the training operation. Following offline training, a single-shot programming pulse was used to update the synaptic weights on the hardware in terms of FeFET channel conductance values. The synaptic weights were normalized between the minimum value (W_{min}) of -1 and the maximum value ($W_{\max}$) of 1. The I_{OFF} of the FeFETs was mapped to $W_{\min}$, and the I_{ON} was mapped to $W_{\max}$. The FeFET-based synaptic core, shown in

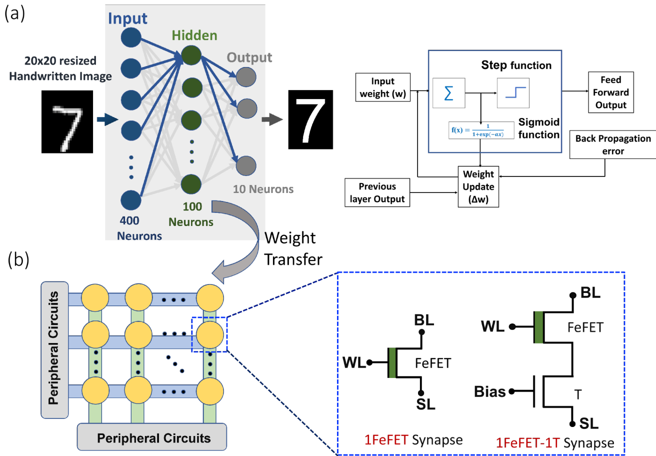

Fig. 5. Schematic representation of the neural network architecture with FeFET-based synaptic core.

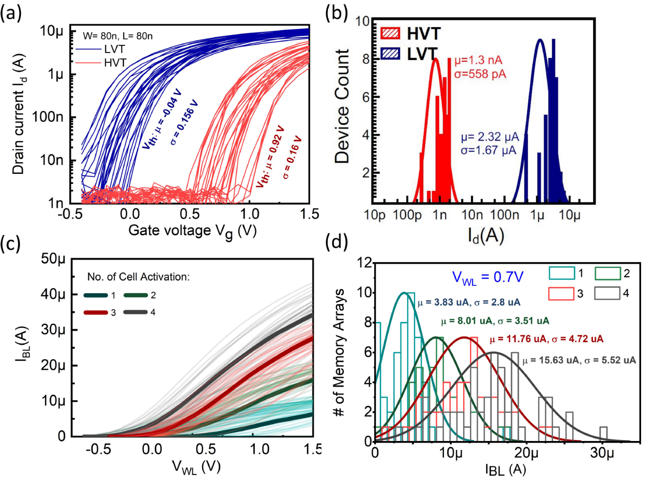

Fig. 6. (a) Transfer characteristics of fabricated devices after WRITE operation with 500 ns pulses. (b) The PDF of the I_d for HVT and LVT states at 0.7 V read voltage. (c) The bit line current (I_{BL}) vs. word line voltage (V_{WL}) for different columns of the crossbar memory array. (d) The PDFs of the I_{BL} show overlapping currents, which makes the quantization process difficult.

Fig. 5, is used to carry out the VMM operation. The output of the VMM is directly digitized using a current-to-digital converter.

III. Results and Discussion

Fig. 6(a) shows the READ operation conducted two seconds after WRITE by applying a voltage ramp with a step size of 100 mV. The READ current's probability density function (PDF) is shown in Fig. 6(b). We observe that drain current device-to-device variation is more at the LVT than at the HVT state. The experimentally measured READ-WRITE of a single FeFET was calibrated with a comprehensive model [58] for the FeFETs to evaluate the characteristics of a memory array. The mean and standard deviation of V_{th} for HVT and LVT states were used for statistical variation analysis of the devices. Fig. 6(c) shows the current ensemble through the bit-lines of different memory array columns according to the voltage applied to the word line (V_{WL}). I_{BL} increases as the number of activation cells increases in a column, demonstrating the multiply and accumulation operation.

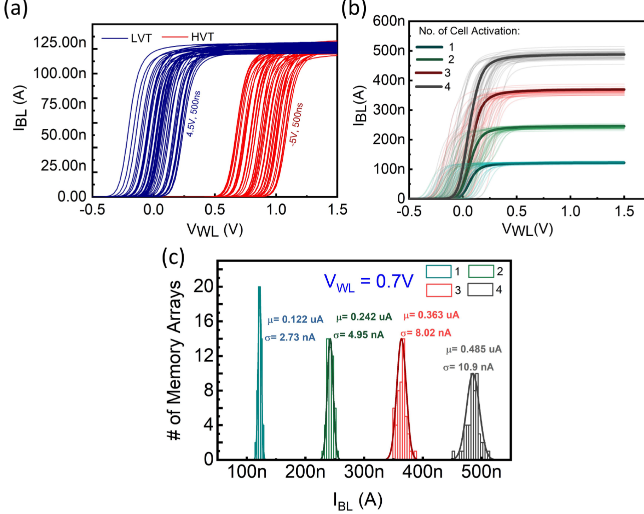

Fig. 7. (a) Transfer characteristics of the 1F-1T synaptic cell after write operation with 500 ns pulses. (b) Simulation of multiply-and-accumulate (MAC) operation on 1F-1T array shows non-overlapping I_{BL} characteristics. (c) The PDF of improved I_{BL}.

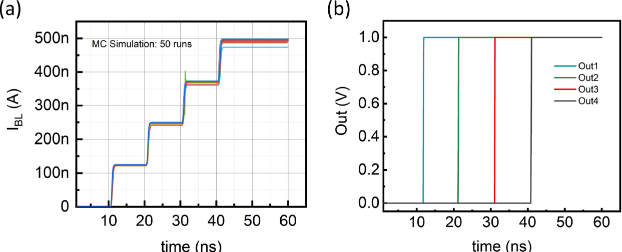

Fig. 8. (a) Transient response I_{BL} upon sequential activation of the LVT FeFETs in the column after each 10ns. (b) The output waveform of sensed output voltage.

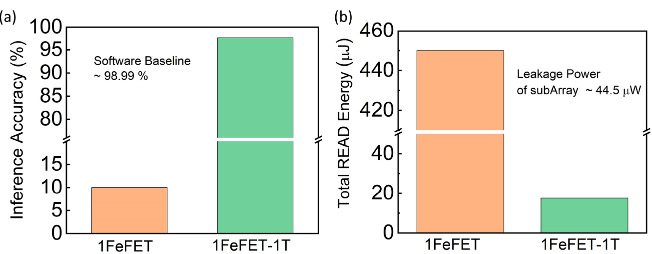

Fig. 9. (a) Inference accuracy and (b) READ energy obtained via system-level simulation of neural networks corroborates the superiority of 1F-1T synapses over 1F synapses.

The PDFs of I_{BL} show that as the number of activation cells increases in the column, the distribution curves start to overlap, making the multiply-and-accumulate (MAC) operation futile for in-memory computing applications as shown in Fig. 6(d). Next, we propose an alternative synaptic cell consisting of a ferroelectric memory transistor cascaded with another logic transistor (1F-1T) to circumvent this issue. Since the variations mainly affect the LVT state, we can efficiently address the problem by limiting the ON current. The logic transistor acts as a current limiter, and the ON current of the FeFET is limited by tuning the gate-source voltage of the cascaded transistor. FeFET switches the state of the cell, and the cascaded transistor controls the cell's ON current. The 1F-1T arrangement thereby makes the ON state current variation independent of the FeFET V_{th} variation and considerably decreases it. The 1F-1T structure was adopted by connecting the FeFET model and the BSIM-4 model [58], [59]. The variation statistics obtained from experimental data were used to evaluate the impact of the current limiting transistor on the variation of I_{BL} through Monte Carlo (MC) simulation for 50 iterations. The variation for the logic transistor cells was not considered assuming they are insignificant compared to the FeFET devices. Fig. 7(a) shows the transfer characteristics of the proposed cell with the same pulsing scheme as a single FeFET cell. It is observed that the ratio of drain currents at HVT and LVT is low compared to the 1F cell. Next, the simulation of the 4 × 4 array with 1F-1T memory devices was conducted. Fig. 7(b) shows non-overlapping MAC operation among different bit-line current levels. Therefore, integrating the current limiter transistor with each FeFET device reduces the variation significantly in the I_{BL}. The PDF of the MAC operation with 1F-1T structures in Fig. 7(c) shows non-overlapping distinguishable PDFs of I_{BL} among different columns with four cells in each column. Transient analysis of the sensed output voltage of the 4-stage CSA shown in Fig. 4 has been carried out, keeping the device-to-device variations of the FeFET in the 1F-1T memory array. The accumulated I_{BL} upon sequential activation of the LVT FeFETs in the column after each 10ns for the 50 iterations of MC simulation is shown in Fig. 8(a). Four reference current sources I_{REF1}, I_{REF2}, I_{REF3} and I_{REF4} has been considered with respective currents of 100 μA, 200 μA, 300 μA and 400 μA. Once the I_{BL} grows beyond I_{REF}, the Output voltage (Out) of the differential comparator goes high. Otherwise, it remains low, as shown in Fig. 8(b). Thus, the sense amplifier can successfully detect the different bit-line current levels of the 1F-1T memory array despite the device-to-device variation of FeFET devices due to the cascaded transistor, which limits the I_{BL} and makes the non-overlapping current levels.

Fig. 9(a) shows the inference accuracy achieved for offline training. 1F-1T exhibits a clear accuracy advantage compared to 1F-based synapses. It achieves 97.6% accuracy, whereas the software benchmark is 98.99%. Further, it is observed that total leakage power is almost the same for both subarrays because the OFF current (I_{off}) is the same for both cells. But total read energy is less for the 1F-1T-based synaptic core than the 1F-based core due to the lower ON current of FeFET. The total read energy of 1F and 1F-1T based cores are 450 μJ and 17.6 μJ, respectively, shown in Fig. 9(b). So, the MLP-based NN with our proposed 1F-1T synaptic core offers excellent immunity towards device variations, accuracy, and low energy consumption. Finally, a detailed comparison of 1F-1T with other emerging memories investigated for in-memory computing applications is shown in Table II. As the table shows, all other emerging memories have several advantages. Still, they are all plagued by problems with device-to-device variability, which renders the VMM operation ineffective. On the other hand, even when FeFET devices are

TABLE II
KEY DEVICE PARAMETERS AND PERFORMANCE

	PCM [51]–[53]	**RRAM** [54]–[56]	**FeRAM** [3], [57]	**FeFET** [30], [31]	**Fe-FinFET** [29]	**1F-1T** **[This Work]**
Multi-bit capacity	Yes	Yes	No	Yes	Yes	Yes
Non-destructive read	Yes	Yes	No	Yes	Yes	Yes
Write voltage	< 3 V	< 3 V	< 2 V	< 5 V	3.5 V	< 5 V
Write energy (J/bit)	18 pJ	$\sim$1 pJ	$\sim$0.1 pJ	1-10 fJ	-	1-10 fJ
Read speed	50 ns	< 10 ns	< 10 ns	< 10 ns	< 50 ns	< 10 ns
Write speed	150 ns	< 10 ns	< 10 ns	< 10 ns	$\sim$100 ns	< 10 ns
Endurance	$> 10^9$	$10^6 - 10^{12}$	$> 10^{12}$	$10^5 - 10^9$	$> 10^{11}$	$10^5 - 10^9$
I_{BL} **Variabilty**	Issue(drift)	Issue(drift)	High(small size)	High(small size)	Low	Very Low

scaled down, our suggested 1F-1T-based in-memory computing architecture reduces the bit-line current variation.

IV. CONCLUSION

In this work, we have proposed a novel 1F-1T memory cell for in-memory computing applications. The operations are evaluated by characterizing the 28 nm HKMG FeFET arrays. The cascaded 1T cell acts as a current limiter and reduces the variation in bit-line current compared to 1FeFET cell-based memory array. Despite FeFET V_{th} variation, we have shown that the current sensing amplifier can accurately digitize the different bit-line current levels of the 1F-1T memory array. We have also demonstrated the impact of device variation on system-level performance in the inference of an MLP-NN. The 1F-1T array-level simulation depicts that such an analog weight cell-based crossbar array accelerates the system-level performance for building in-memory-computing hardware.

ACKNOWLEDGMENT

BMWk and the State of Saxony in the frame of IPCEI.

REFERENCES

[1] D. Ielmini and H.-S. P. Wong, "In-memory computing with resistive switching devices," *Nature Electron.*, vol. 1, no. 6, pp. 333–343, 2018.

[2] A. Sebastian, M. Le Gallo, R. Khaddam-Aljameh, and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nature Nanotechnol.*, vol. 15, no. 7, pp. 529–544, 2020.

[3] A. I. Khan, A. Keshavarzi, and S. Datta, "The future of ferroelectric field-effect transistor technology," *Nature Electron.*, vol. 3, no. 10, pp. 588–597, 2020.

[4] S. Jung et al., "A crossbar array of magnetoresistive memory devices for in-memory computing," *Nature*, vol. 601, no. 7892, pp. 211–216, 2022.

[5] V. Parmar et al., "Demonstration of differential mode FeFET-array-based IMC-macro for realizing multi-precision mixed-signal ai accelerator," *Adv. Intell. Syst.*, vol. 5, 2023, Art. no. 2200389.

[6] A. Kazemi et al., "FeFET multi-bit content-addressable memories for in-memory nearest neighbor search," *IEEE Trans. Comput.*, vol. 71, no. 10, pp. 2565–2576, Oct. 2022, doi: 10.1109/TC.2021.3136576.

[7] F. M. Bayat, M. Prezioso, B. Chakrabarti, H. Nili, I. Kataeva, and D. Strukov, "Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits," *Nature Commun.*, vol. 9, no. 1, pp. 1–7, 2018.

[8] B. Choi et al., "Resistive switching mechanism of TiO2 thin films grown by atomic-layer deposition," *J. Appl. Phys.*, vol. 98, no. 3, 2005, Art. no. 033715.

[9] E. Linn, R. Rosezin, C. Kügeler, and R. Waser, "Complementary resistive switches for passive nanocrossbar memories," *Nature Mater.*, vol. 9, no. 5, pp. 403–406, 2010.

[10] S. Pande, B. Chakrabarti, and A. Chakravorty, "Thermal crosstalk analysis in RRAM passive crossbar arrays," 2023, arXiv:2304.01439.

[11] S. Pande, S. Balanethiram, B. Chakrabarti, and A. Chakravorty, "A physics-based compact model of thermal resistance in RRAMs," *Solid-State Electron.*, vol. 204, 2023, Art. no. 108636.

[12] P. Basnet, E. C. Anderson, F. F. Athena, B. Chakrabarti, M. P. West, and E. M. Vogel, "Asymmetric resistive switching of bilayer HfO_x/ALO_y and ALO_y/HfO_x memristors: The oxide layer characteristics and performance optimization for digital set and analog reset switching," *ACS Appl. Electron. Mater.*, vol. 5, no. 3, pp. 1859–1865, 2023.

[13] W. Haensch et al., "Compute in-memory with non-volatile elements for neural networks: A review from a co-design perspective," *Adv. Materials*, 2022, Art. no. 2204944.

[14] A. Sebastian, M. Le Gallo, G. W. Burr, S. Kim, M. BrightSky, and E. Eleftheriou, "Tutorial: Brain-inspired computing using phase-change memory devices," *J. Appl. Phys.*, vol. 124, no. 11, 2018, Art. no. 111101.

[15] S. De, W.-X. Bu, B.-H. Qiu, C.-J. Su, Y.-J. Lee, and D. D. Lu, "Alleviation of charge trapping and flicker noise in HfZrO2-based ferroelectric capacitors by thermal engineering," in *Proc. Int. Symp. VLSI Technol. Syst. Appl.*, 2021, pp. 1–2.

[16] M. Jerry et al., "Ferroelectric FET analog synapse for acceleration of deep neural network training," in *Proc. IEEE Int. Electron Devices Meeting*, 2017, pp. 6.2.1–6.2.4.

[17] S. De et al., "Robust binary neural network operation from 233k to 398k via gate stack and bias optimization of ferroelectric FinFET synapses," *IEEE Electron Device Lett.*, vol. 42, no. 8, pp. 1144–1147, Aug. 2021.

[18] S. De et al., "Neuromorphic computing with deeply scaled ferroelectric FinFET in presence of process variation, device aging and flicker noise," 2021, arXiv:2103.13302.

[19] S. De et al., "Random and systematic variation in nanoscale Hf0. 5Zr0. 5O2 ferroelectric FinFETs: Physical origin and neuromorphic circuit implications," *Front. Nanotechnol.*, vol. 3, 2022.

[20] H. Mulaosmanovic et al., "Novel ferroelectric FET based synapse for neuromorphic systems," in *Proc. Symp. VLSI Technol.*, 2017, pp. T176–T177.

[21] S. De et al., "28nm HKMG based current limited FeFET crossbar-array for inference application," *IEEE Trans. Electron Device*, vol. 69, no. 12, pp. 7194–7198, Dec. 2022.

[22] S. De et al., "Roadmap of ferroelectric memories: From discovery to 3D integration," Jun. 2023, doi: 10.36227/techrxiv.23573523.v1.

[23] X. Ma et al., "A 2-transistor-2-capacitor ferroelectric edge compute-in-memory scheme with disturb-free inference and high endurance," *IEEE Electron Device Lett.*, vol. 44, no. 7, pp. 1088–1091, Jul. 2023.

[24] V. Parmar et al., "Demonstration of differential mode FeFET-array for multi-precision storage and IMC applications," in *Proc. Int. Symp. VLSI Technol. Syst. Appl.*, 2023, pp. 1–2.

[25] J.-H. Hsuen et al., "Demonstration of large polarization in si-doped HfO2 metal–ferroelectric–insulator–semiconductor capacitors with good endurance and retention," in *Proc. Int. Symp. VLSI Technol. Syst. Appl.*, 2023, pp. 1–2.

[26] S. De et al., "Demonstration of multiply-accumulate operation with 28nm FeFET crossbar array," *IEEE Electron Device Lett.*, vol. 43, no. 12, pp. 2081–2084, Dec. 2022.

[27] S. De, "First demonstration of ultra-high precision 4kb 28nm HKMG 1FeFET-1T based memory array macro for highly scaled deep learning applications," 2022, doi: 10.36227/techrxiv.19491212.v1.

[28] S. Dünkel et al., “A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond,” in *Proc. IEEE Int. Electron Devices Meeting*, 2017, pp. 19.7.1–19.7.4.

[29] S. De et al., “Ultra-low power robust 3 bit/cell Hf 0.5 Zr 0.5 O 2 ferroelectric FinFET with high endurance for advanced computing-in-memory technology,” in *Proc. Symp. VLSI Technol.*, 2021, pp. 1–2.

[30] S. Beyer et al., “FeFET: A versatile CMOS compatible device with game-changing potential,” in *Proc. IEEE Int. Memory Workshop*, 2020, pp. 1–4.

[31] M. Lederer et al., “Ferroelectric field effect transistors as a synapse for neuromorphic application,” *IEEE Trans. Electron Devices*, vol. 68, no. 5, pp. 2295–2300, May 2021.

[32] S. De et al., “READ-optimized 28nm HKMG multibit FeFET synapses for inference-engine applications,” *IEEE J. Electron Devices Soc.*, vol. 10, pp. 637–641, 2022.

[33] S. De, M. A. Baig, B.-H. Qiu, H.-H. Le, Y.-J. Lee, and D. Lu, “Neuromorphic computing with Fe-FinFETs in the presence of variation,” in *Proc. Int. Symp. VLSI Technol. Syst. Appl.*, 2022, pp. 1–2.

[34] T. Ali et al., “Study of nanosecond laser annealing on silicon doped hafnium oxide film crystallization and capacitor reliability,” in *Proc. IEEE Int. Memory Workshop*, 2022, pp. 1–4.

[35] F. Müller et al., “Multilevel operation of ferroelectric FET memory arrays considering current percolation paths impacting switching behavior,” *IEEE Electron Device Lett.*, vol. 44, no. 5, pp. 757–760, May 2023.

[36] S. De et al., “Tri-gate ferroelectric FET characterization and modelling for online training of neural networks at room temperature and 233k,” in *Proc. IEEE Device Res. Conf.*, 2020, pp. 1–2.

[37] S. De, M. Lederer, Y. Raffel, F. Müller, K. Seidel, and T. Kämpfe, “Roadmap for ferroelectric memory: Challenges and opportunities for IMC applications,” in *Proc. IEEE 19th Int. SoC Des. Conf.*, 2022, pp. 167–168.

[38] Y. Raffel et al., “28nm high-k-metal gate ferroelectric field effect transistors based synapses–A comprehensive overview,” *Memories-Materials Devices, Circuits Syst.*, vol. 4, 2023, Art. no. 100048.

[39] M. N. K. Alam et al., “On the characterization and separation of trapping and ferroelectric behavior in HfZrO FET,” *IEEE J. Electron Devices Soc.*, vol. 7, pp. 855–862, 2019.

[40] G. Bersuker et al., “Mechanism of electron trapping and characteristics of traps in hfO_2 gate stacks,” *IEEE Trans. Device Mater. Rel.*, vol. 7, no. 1, pp. 138–145, Mar. 2007.

[41] K. Ni, W. Chakraborty, J. Smith, B. Grisafe, and S. Datta, “Fundamental understanding and control of device-to-device variation in deeply scaled ferroelectric FETs,” in *Proc. IEEE Symp. VLSI Technol.*, 2019, pp. T40–T41.

[42] S. De et al., “Uniform crystal formation and electrical variability reduction in hafnium-oxide-based ferroelectric memory by thermal engineering,” *ACS Appl. Electron. Mater.*, vol. 3, no. 2, pp. 619–628, 2021.

[43] T. Ali et al., “Impact of the ferroelectric stack lamination in si doped hafnium oxide (HSO) and hafnium zirconium oxide (HZO) based FeFETs: Toward high-density multi-level cell and synaptic storage,” *Electron. Mater.*, vol. 2, no. 3, pp. 344–369, 2021.

[44] Y. Raffel et al., “Endurance improvements and defect characterization in ferroelectric FETs through interface fluorination,” in *Proc. IEEE Int. Memory Workshop*, 2022, pp. 1–4.

[45] A. J. Tan et al., “Ferroelectric HfO_2 memory transistors with high-κ interfacial layer and write endurance exceeding 10^{10} cycles,” *IEEE Electron Device Lett.*, vol. 42, no. 7, pp. 994–997, Jul. 2021.

[46] Y. Raffel et al., “Interfacial layer engineering to enhance noise immunity of FeFETs for IMC applications,” in *Proc.IEEE Int. Conf. IC Des. Technol.*, 2022, pp. 8–11.

[47] S. De, W.-X. Bu, B.-H. Qiu, C.-J. Su, Y.-J. Lee, and D. D. Lu, “Alleviation of charge trapping and flicker noise in HfZrO2-based ferroelectric capacitors by thermal engineering,” in *Proc.IEEE Int. Symp. VLSI Technol. Syst. Appl.*, 2021, pp. 1–2.

[48] M.-K. Seo et al., “A 130-nm 0.9-v 66-MHz 8-Mb (256K/spl times/32) local SONOS embedded flash EEPROM,” *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 877–883, Apr. 2005.

[49] A. Conte, G. L. Giudice, G. Palumbo, and A. Signorello, “A high-performance very low-voltage current sense amplifier for nonvolatile memories,” *IEEE J. Solid-State Circuits*, vol. 40, no. 2, pp. 507–514, Feb. 2005.

[50] P.-Y. Chen, X. Peng, and S. Yu, “NeuroSim+: An integrated device-to-algorithm framework for benchmarking synaptic devices and array architectures,” in *Proc. IEEE Int. Electron Devices Meeting*, 2017, pp. 6.1.1–6.1.4.

[51] Y. Choi et al., “A 20nm 1.8 v 8Gb PRAM with 40MB/s program bandwidth,” in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2012, pp. 46–48.

[52] K.-J. Lee et al., “A 90nm 1.8 V 512 Mb diode-switch PRAM with 266 Mb/s read throughput,” *IEEE J. Solid-State Circuits*, vol. 43, no. 1, pp. 150–162, Jan. 2008.

[53] B. C. Lee, E. Ipek, O. Mutlu, and D. Burger, “Architecting phase change memory as a scalable dram alternative,” in *Proc. 36th Annu. Int. Symp. Comput. Architecture*, 2009, pp. 2–13.

[54] S. Yu and P.-Y. Chen, “Emerging memory technologies: Recent trends and prospects,” *IEEE Solid-State Circuits Mag.*, vol. 8, no. 2, pp. 43–56, Spring 2016.

[55] D. Ielmini, “Resistive switching memories based on metal oxides: Mechanisms, reliability and scaling,” *Semicond. Sci. Technol.*, vol. 31, no. 6, 2016, Art. no. 063002.

[56] S.-S. Sheu et al., “A 4Mb embedded SLC resistive-RAM macro with 7.2ns read-write random-access time and 160ns MLC-access capability,” in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2011, pp. 200–202.

[57] S. Khanna, S. C. Bartling, M. Clinton, S. Summerfelt, J. A. Rodriguez, and H. P. McAdams, “An FRAM-based nonvolatile logic MCU SoC exhibiting 100% digital state retention at VDD$=$0 V achieving zero leakage with $<$ 400-ns wakeup time for ULP applications,” *IEEE J. Solid-State Circuits*, vol. 49, no. 1, pp. 95–106, Jan. 2014.

[58] S. Deng et al., “A comprehensive model for ferroelectric FET capturing the key behaviors: Scalability, variation, stochasticity, and accumulation,” in *Proc. IEEE Symp. VLSI Technol.*, 2020, pp. 1–2.

[59] X. Xi et al., “BSIM4.3.0 MOSFET model-user’s manual,” 2003. [Online]. Available: http://www-device.eecs.berkeley.edu/bsim3/bsim4.html