Performance-Reliability Co-Optimized IGZO 3T0C-based Analog CIM with Excellent Synaptic Properties and Fully-Parallel Update

Junyoung Choi[1], Cheng-Wei Cheng[2], Ga-eun Heo[3], Bongjun Ju[3], Byoungwoo Lee[4], Jimin Lee[1], Jinho Byun[1], Jay Chey[2], Paul Solomon[2], Minseong Um[5], Minil Kang[6], Hyung-Min Lee[5], Martin M. Frank[2*], Seyoung Kim[1,3*]
[1]Dept. of Elec. Eng., POSTECH, Pohang, Korea, [2]IBM T. J Watson Research Center, NY, USA,
*E-mail: kimseyoung@postech.ac.kr, mmfrank@us.ibm.com
[3]Graduate School of Semicon. Tech., POSTECH, Pohang, Korea, [4]Dept of Mat. Sci. & Eng., POSTECH, Pohang, Korea,
[5]School of Elec. Eng., [6]Dept. of Semicon. Sys. Eng., Korea University, Seoul, Korea

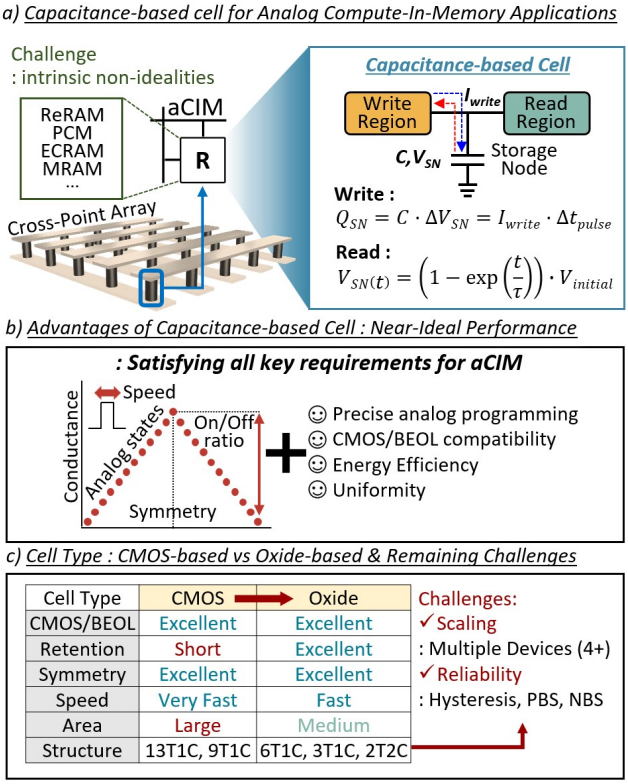
***Abstract*: We demonstrate a performance–reliability co-optimized (PRCO) IGZO-based 3T0C synaptic cell for analog compute-in-memory (aCIM) applications. By employing passivated novel IGZO transistor with small hysteresis and robust PBS/NBS stability, our 3T0C cell simultaneously achieves key synaptic requirements for aCIM, including ultrafast switching (10ns), high symmetry, multi-state programmability (>4bit). The dual-transistor write scheme of 3T0C enables fully parallel and highly selective update (>10^3), confirming reliable weight update without interference in cross-point array (CPA) architecture. Based on superior and reliable properties, we experimentally fabricate a 2K IGZO 3T0C CPA and validate both simulation and experimental demonstration, showing that 3T0C characteristics can achieve high training accuracy across diverse neural networks and learning algorithms including transformer-based mobileBERT.**

***Keywords—Analog* Compute-In-Memory, 3T0C, Cross-Point-Array, IGZO-based Synaptic Memory**

I. Introduction

The rapid growth of deep neural network (DNN) applications has driven the demand for highly efficient hardware accelerators, and analog compute-in-memory (aCIM) has emerged as a promising solution to overcome the memory–compute bottleneck by performing matrix–vector multiplication directly within memory arrays [1]. As an aCIM architecture, resistive cross-point array, where non-volatile memory (NVM) devices such as ReRAM, PCM, and ECRAM serve as analog synaptic elements, has been widely explored [2]. However, despite their advantages, these NVM-based devices suffer from intrinsic non-idealities, including asymmetric weight updates, stochastic switching behavior, device-to-device variability, which degrade training accuracy and reliability. To address these challenges, capacitance-based synaptic cells have been proposed, offering excellent switching properties, improved controllability and CMOS compatibility [3,4].

Nevertheless, conventional capacitance-based synaptic cell designs relying on multiple transistor(s) and capacitor(s) introduces structural complexity, scaling challenges and retention degradation due to off-leakage of CMOS transistors. To address these issues, oxide transistors such as IGZO with extremely low off-leakage were adopted and realized the scaled cell such as 6T1C, 3T1C, 2T2C [5-7] and the most-compact 3T0C cell [8], which we previously reported for scalable and efficient aCIM. Nevertheless, intrinsic defects in IGZO and its interfaces act as electron traps, resulting reliability challenges. In this work, we suggest passivation induced performance-reliability co-optimized 3T0C synaptic cell and demonstrate its 2K array-level operation, validating its strong potential for scalable and reliable aCIM applications.



Cell Type	CMOS → Oxide	
CMOS/BEOL	Excellent	Excellent
Retention	Short	Excellent
Symmetry	Excellent	Excellent
Speed	Very Fast	Fast
Area	Large	Medium
Structure	13T1C, 9T1C	6T1C, 3T1C, 2T2C

Fig. 1. a) Challenges of conventional NVMs and proposed capacitance-based cell for aCIM applications. b) Advantages of capacitance-based cell showing ideal performances. c) Oxide-based capacitor cells for overcoming challenges derived from CMOS-based capacitor cells, and remaining limitations such as scaling and reliability.

II. Highly Reliable IGZO Transistor Characteristics

In Fig. 2a, the PRCO-IGZO 3T0C cell modulates conductance by controlling charge at the storage node through dual-gate write transistors (N_1, N_2) and sensing through read transistor (N_3). IGZO 3T0C cells were fabricated using optical lithography (Fig. 2b), employing optimized encapsulation and thermal processing [9-10]. IGZO channels were sputtered from $InGaZnO_4$ target onto O_3-reacted ALD HfO_2 gate dielectrics and W bottom gate electrodes to avoid hydrogen-induced damage.

*The authors recognize the Microelectronics Research Laboratory (MRL) at IBM Research for assistance with device fabrication. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-09322969) and K-CHIPS (Korea Collaborative & High-tech Initiate for Prospective Semiconductor Research) (2410010977, RS-2023-00235402, 23003-15FC) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). J. Choi acknowledges the financial support by the Hyundai Motor Chung Mong-Koo Foundation through a scholarship.

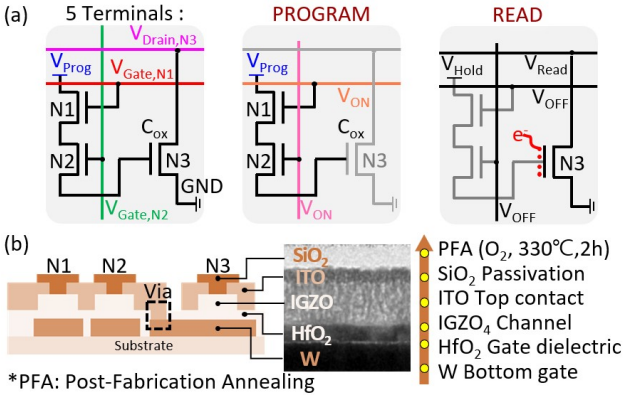


Fig. 2. (a) Definition of 5 terminals in 3T0C cell: V_{Prog}, $V_{Gate,N1}$, $V_{Gate,N2}$, $V_{Drain,N3}$, GND and operation mechanism of 3T0C cell. (b) Fabrication flow of 3T0C cell and TEM image of PRCO-IGZO transistor.

Id-Vg Curve of Performance-Reliability Co-Optimized IGZO Transistor

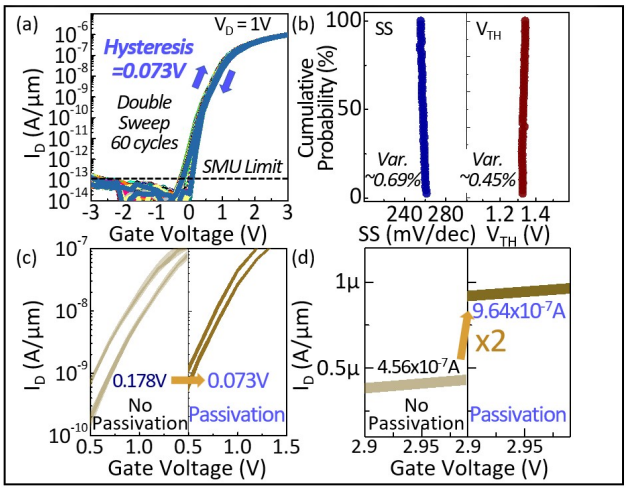


Fig. 3. (a) Repeated I_D-V_G curves of PRCO-IGZO transistor with small hysteresis and (b) extremely low variation of SS and V_{TH}. Passivation improves performance-reliability simultaneously, decreasing hysteresis and increasing on-current.

Subsequently, optimized condition of ITO top contact, SiO_2 passivation, and O_2 annealing were performed to simultaneously improve reliability and performance of 3T0C cells. Fig. 3a shows repeated I_D–V_G characteristics over 60 double-sweep cycles, exhibiting a small hysteresis of 73 mV. The extracted SS and V_{TH} variations are extremely low (Fig. 3b). Fig. 3c,d compares sweep characteristics of IGZO transistor with and without SiO_2 passivation. The passivated device exhibits higher on-current, improved SS, and a smaller ΔV_{TH}, directly indicating the effects of passivation in improving performance and reliability simultaneously. Moreover, I_D-V_D curve of and the overdrive-voltage dependence of the ΔV_{TH} are evaluated in Fig. 4(a,b), indicating that our IGZO transistors are robust to bias stress, enabling accurate and stress-free 3T0C operation.

III. Superior Performance of PRCO-IGZO 3T0C Cell

Based on the measurement schemes of 3T0C presented in Fig. 5, our 3T0C simultaneously achieves all key synaptic properties including acceptable symmetry, large on/off ratio with 50ns T_{pulse} (Fig. 5a). Moreover, 3T0C shows crosstalk-free multi-state (>4bit) with small V_{SN} shift (<50mV) for 300s after write operation by implementing low off-leakage (6.12 aA/μm) PRCO-IGZO transistor. (Fig. 5b) During read, -1V was applied to both gates of the dual-gate write transistors (N_1 & N_2) to suppress the leakage current.

Accurate and Stress-robust programming for 3T0C Operation

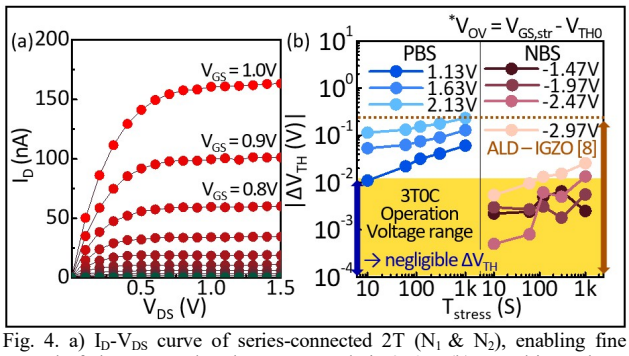


Fig. 4. a) I_D-V_{DS} curve of series-connected 2T (N_1 & N_2), enabling fine control of charge stored at the storage node in 3T0C. (b) Overdrive-voltage dependence of V_{TH} shift of PRCO-IGZO transistor, confirming PBS/NBS-robust 3T0C operation with negligible V_{TH} shift stabilizing analog programming.

Near-Ideal Performances compared to Conventional NVMs

Terminal	100 UP/DOWN		Retention		Endurance	
	Voltage	width	Voltage	width	Voltage	width
$V_{Gate, N1, N2}$	3.5V (Up) 2V (Down) -1V(OFF)	50ns	-1V	Const.	1.5V (Up) 0.2V (Down) -0.7V(OFF)	500ns
V_{Prog}	4V (Up) 0V (Down)	Const.	0V	Const.	1.5V (Up) -0.2V (Down)	Const.
$V_{Drain, N3}$	0.1V	Const.	0.1V	Const.	0.1V	Const.

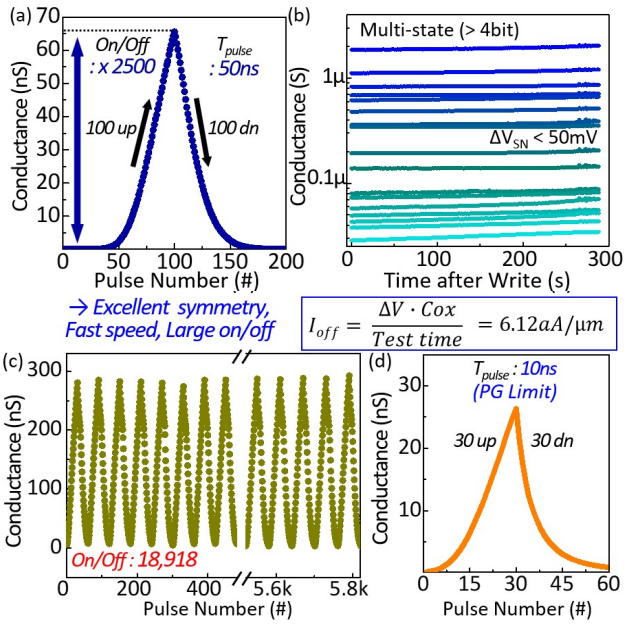


Fig. 5. Summary of characterization condition of 3T0C cell. (a) Acceptable symmetry, large on/off ratio (>2,500) switching of 3T0C cell with 100 up/down pulses with 50ns T_{pulse}. (b) Crosstalk-free, stable multi-state retention (>4bit) with small V_{SN} shift (<50mV). (c) Endurance over $6x10^3$ pulses with low variation of G_{max}, G_{min}. (d) Ultrafast 10ns switching enabling fast and energy-efficient operation for aCIM.

In particular, when both N_1 and N_2 are turned off, two high-resistance channels are connected in series, further enhancing retention characteristics. In addition, the 3T0C device shows stable endurance over $6x10^3$ pulses with negligible degradation (Fig. 5c), and ultra-fast switching down to 10ns (Fig. 5d), enabling energy-efficient aCIM applications. These results indicate that 3T0C provides enhanced overall performance compared to conventional NVM devices.

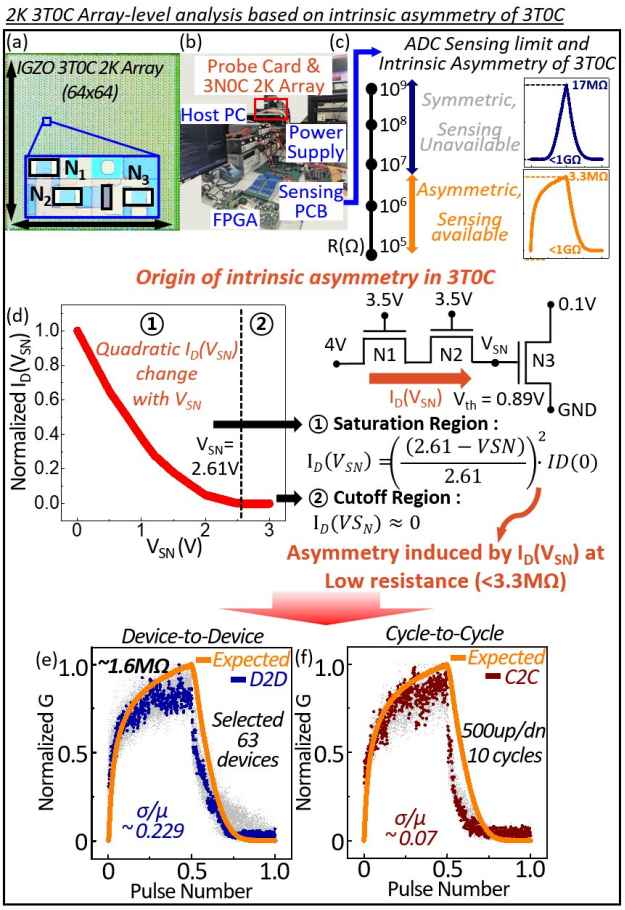


Fig. 6. (a) Firstly demonstrated 3T0C 2K Array and (b) Measurement setup. (c) Intrinsic asymmetry in 3T0C induced by ADC sensing limit and (d) analysis on intrinsic asymmetry in 3T0C due to quadratic I_D change with V_{SN}. (e) Device-to-device variation of selected 63 devices in 2K array and (f) cycle-to-cycle variation for 10 cycles. Low-resistance operation (due to ADC sensing limit) induces asymmetry in 3T0C as expected.

IV. First IGZO 3T0C 2K Array Demonstration

Fig. 6a,b present the firstly demonstrated PRCO-IGZO-based 3T0C 2K cross-point array and measurement setup. In Fig. 6c, the 3T0C cell exhibits intrinsic asymmetry depending on the operating conductance range: acceptable symmetry is achieved in the relatively high-resistance regime (~17Mohm), whereas asymmetry emerges at relatively low-resistance (~3.3Mohm). The resistance at which asymmetry begins to emerge can be tuned by the applied V_{Prog}, $V_{Gate,N1}$, $V_{Gate,N2}$ of the IGZO dual-gate transistors as well as the on-current of the IGZO read transistor. The origin of the intrinsic asymmetry is analyzed in Fig. 6d. During programming, the dual-gate write transistors operate under fixed bias conditions, and as V_{SN} increases, both V_{GS} and V_{DS} decrease. Since the write transistors operate in the saturation region, the drain current is primarily governed by V_{GS}, regardless of V_{DS}, resulting in a quadratic decrease of current with increasing V_{SN}. Consequently, the amount of charge delivered to the storage node decreases, leading to asymmetric weight updates. Based on the extracted V_{TH} (0.89V) from Fig. 3a, V_{SN} can increase up to 2.61 V, beyond which N_1 and N_2 enter the cutoff region and no further current flows, defining the minimum resistance state. Fig. 6e,f demonstrate the switching behavior of 3T0C cells within the 2K array, confirming their conductance modulation as expected. While single 3T0C cell measurements (Fig. 5a-d) using a Keithley 2636B allow access to the full resistance

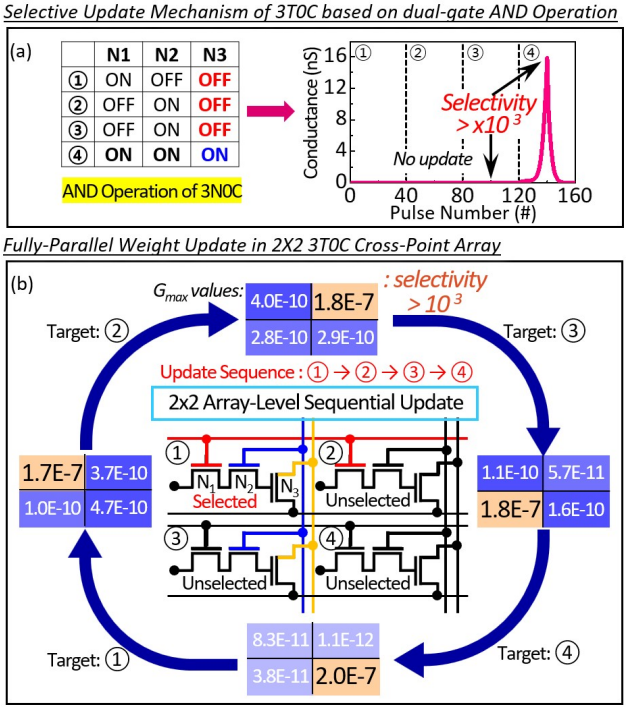


Fig. 7. (a) Selective update in single 3T0C cell enabled by dual-gate control and (b) Sequential weight update in 2x2 3T0C cross-point array with high selectivity (>10^3)

range, 2K array measurements with a customized PCB are constrained by the ADC sensing limit, requiring operation in the low-resistance region where asymmetry becomes significant. In this regime, the observed readout noise during update originates from the mismatch between the MPU and ADC in the peripheral PCB rather than from the 3T0C cells themselves. Despite this, Fig. 6e,f demonstrates device-to-device and cycle-to-cycle variation of 0.229 and 0.07, which is highly acceptable with respect to the aCIM specification [11]. Fig. 7a illustrates selective update enabled by the dual-gate structure of the 3T0C cell, where programming occurs only when both N_1 and N_2 are simultaneously ON, while no update is induced if either N_1 or N_2 is off. Fig. 7b further demonstrates highly selective sequential weight updates in a 2x2 sub-array, achieving a selectivity exceeding 10^3 between selected and unselected cells. This confirms fully parallel update capability without cell-to-cell interference, which is particularly suitable for stochastic pulse update schemes [11].

V. Device-Algorithm Co-Optimization for AI Training

Fig. 8a compares the conventional stochastic gradient descent (SGD) algorithm with the Tiki-Taka version 1 (TTv1) algorithm used for neural network training [12]. In the conventional SGD algorithm, a single core (C) array is utilized for both inference and weight update. In contrast, the TTv1 algorithm employs both a core (C) array and an auxiliary (A) array: Inference is first performed in the C array, followed by weight updates in the A array. The accumulated gradients in the A array are then transferred back to update the C array. This update scheme effectively mitigates intrinsic non-idealities of cross-point devices, particularly update errors caused by asymmetry, thereby preventing degradation in training accuracy. To systematically analyze effective neural network training strategies leveraging both the superior 3T0C characteristics (Fig. 4) and intrinsic asymmetry (Fig.6), we performed

High Neural Network Training Simulation Accuracy using TTv1 Algorithm

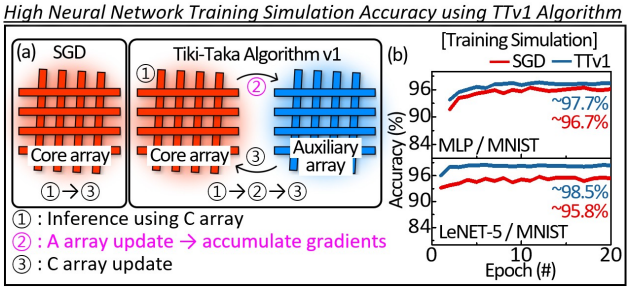


Fig. 8. (a) Comparison of tiki-taka algorithm (TTv1) with SGD and (b) MNIST training simulation accuracy on various neural networks.

First Demonstration of Array-Level Linear-Regression using TTv1 Algorithm

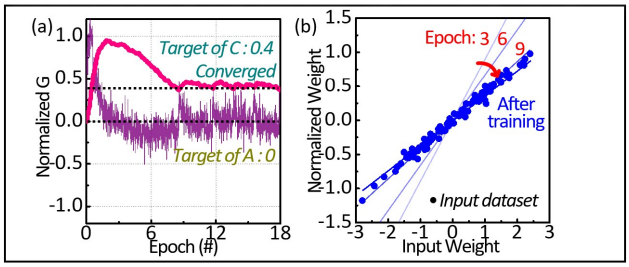


Fig. 9. (a) First experimental demonstration of TTv1 using 3T0C array, showing A/C arrays converging to target values and (b) Linear regression results.

Transformer-based MobileBERT Model Training for Edge-AI Computing

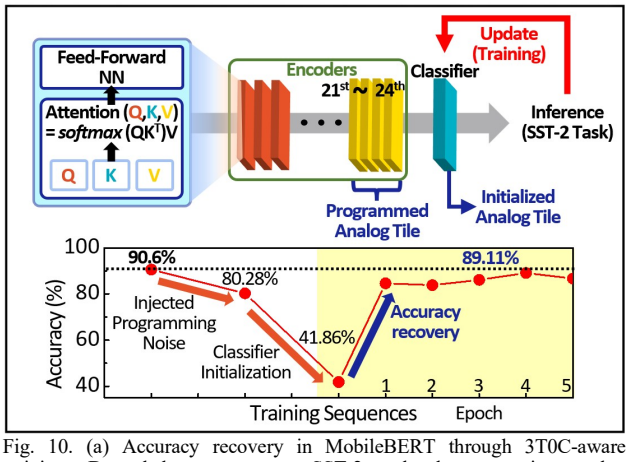


Fig. 10. (a) Accuracy recovery in MobileBERT through 3T0C-aware training. Degraded accuracy on SST-2 task due to naive analog programming and FC initialization was recovered by replacing four encoder layers and FC layer to superior 3T0C-based analog layers.

simulations using both SGD and TTv1 algorithms across various neural network models. Based on the MNIST dataset, the superior characteristics of 3T0C enable high training accuracy even with conventional SGD, with simulated accuracies of 96.7% and 95.8% for multi-layer perceptron (MLP) and LeNet-5 networks, respectively. By further implementing the asymmetry-aware TTv1 algorithm reflecting the behavior in Fig. 6, the accuracy is improved to 97.7% and 98.5%, demonstrating enhanced robustness against device non-idealities. In Fig. 9a, we firstly report array-level on-device linear regression demonstration using the TTv1 algorithm implemented on a 3T0C array. The weights stored in the A and C arrays successfully converge to their respective target values, confirming accurate weight transfer, stable conductance modulation, and successful on-device training capability. This experimental result provides direct validation of the proposed training framework at the array level. To evaluate the applicability of 3T0C to transformer models, MobileBERT model was applied to the SST-2 task, replacing the last four encoder layers and the

TABLE I. Comparison of 3T0C with previous capacitance-based cells

	2018 MWSCAS [3]	2023 Adv. Sci [5]	2025 IEEE EDS [6]	2025 Mater. Horiz. [8]	2025 IEDM [7]	**This work**
Memory Type	CMOS 9T1C	IGZO 6T1C	IGZO 3T1C	IGZO 3T0C	IGZO 2T2C	**IGZO 3T0C**
Multi-Bit	X	X	3 bit (300min)	2 bit (200s)	3 bit (100s)	**> 4 bit (300s)**
Speed, On/Off	50ns, ~12	300ns, ~5	200ns, N/A	10ns ~38	1µs, N/A	**< 10ns ~166**
Array Size	4 x 5	5 x 5	X	X	X	**64 x 32 (2K)**
NN workload	CNN MNIST	CNN, MNIST	MLP, MNIST	MLP, MNIST	CNN, MNIST	**Mobile BERT**

classifier with 3T0C-based analog layers. Using conventional RRAM device models were used, the accuracy dropped significantly from 90.6% to 80.28%, and further to 41.86% after fully-connected (FC) layer initialization. In contrast, by incorporating 3T0C-based analog layers and performing hardware-aware training, the accuracy was successfully recovered to 89.11%, demonstrating robust analog training capability even for deep transformer architectures.

VI. Conclusion

We demonstrated a PRCO-IGZO-based 3T0C synaptic cell and the first 2K cross-point array for analog compute-in-memory. The proposed 3T0C simultaneously achieves ultrafast switching (10 ns), acceptable symmetry, large on/off ratio (>2,500), and stable multi-level operation (>4bit) with minimal V_{SN} shift (<50 mV). As a result, on-device training is successfully achieved, with MNIST accuracies up to 98.5% and successful accuracy recovery to 89.11% in transformer-based MobileBERT. These results highlight the strong potential of 3T0C for high-performance, scalable and efficient analog AI training.

References

[1] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing," *Nat. Mater*, 18, 309-323, 2019.

[2] A. Sebastian, M. L. Gallo, R. Khaddan-Aljameh and E. Eleftheriou, "Memory devices and applications for in-memory computing," *Nat. Nanotechnol*, 15, 529-544, 2020.

[3] S. Kim, T. Gokmen, H. M. Lee and W. E. Haensch, "Analog CMOS-based resistive processing unit for deep neural network training," *MWSCAS, 2017.*

[4] Y. Li et al., "Capacitor-based cross-point array for analog neural network with record symmetry and linearity," in *IEEE Symp. On* VLSI Technology, 2018.

[5] J. Won et al., "Device-algorithm co-optimization for an on-chip trainable capacitor-based synaptic device with IGZO TFT and retention-centric Tiki-Taka algorithm," *Adv. Sci*, 10, 29, 2023.

[6] M. Kang et al., "Enabling selective and tunable weight updates in all-InGaZnO 3-transistor 1-capacitor synaptic circuits for on-chip training," *IEEE J-EDS*, 13, 510-515, 2025.

[7] S. Park et al., "CMOS-compatible oxide TFT-based 2T-2C near-pixel analog compute cell with ADC/DAC-free sensor interface for energy-efficient feature extraction," in 2025 *IEEE International Electron Devices Meeting (IEDM)*, 2025.

[8] J. Choi et al., "Highly-efficient and scalable TrioN (3N0C) synaptic cell for analog process-in-memory," *Mater. Horiz*, 12, 7509-7519, 2025.

[9] C. W. Cheng et al., "Innovative oxide transistor satisfying performance and reliability simultaneously by understanding of physics and materials properties," arXiv:2509.07886, 2025.

[10] C. W. Cheng et al., "Reliable and high performance IGZO and In_2O_3 transistors via channel capping," arXiv:2603.23341, 2026.

[11] T. Gokmen and Y. Vlasov, "Acceleration of deep neural network training with resistive cross-point devices: design considerations," *Front. Neurosci*, 10, 2016.

[12] T. Gokmen and W. Haensch, "Algorithm for training neural networks on resistive device arrays," 14, 2020.