First Demonstration of IGO-IGZO based In-sensor-memory Computing Enables Dual-Stream C-RNN for Software-level Accuracy Spatiotemporal Machine Vision

Shuaidi Zhang[=1,2], Ke Hu[=1,2], Xiaoyi Zhang[1,2], Yao Li[1,2], Zhi Li[1,2], Yanyu Yang[1,2], Xufan Li[1,2], Cailian Ma[1,2], Gaobo Xu[1,2], Jiawei Wang[*1,2], Lingfei Wang[*1,2], Feng Zhang[*1,2], Ling Li[1,2] and Ming Liu[1,2]
[1]State Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of the Chinese Academy of Sciences, Beijing, China; [2]University of Chinese Academy of Sciences, Beijing, China
*email: wangjiawei@ime.ac.cn; wanglingfei@ime.ac.cn; Zhangfeng_ime@ime.ac.com, =equally contributed

***Abstract*—In the in-sensor computing (ISC) paradigm, ultra-high energy efficiency shows great promise for real-time sensor data preprocessing but faces challenges in encoding and processing multi-dimensional optical data like machine vision. To address limitations in sensing modes, non-linearity, and accuracy loss, a synergistic technology-architecture-algorithm (TAA) design is essential. For the first time, we propose an IGO-IGZO-based in-sensor-memory computing (ISMC) architecture that leverages reconfigurable IGO phototransistors with high photoresponsivity (7.3×10^4 mA/W) and robust bias-temperature instability (BTI) performance, as well as 2T0C-IGZO-DRAM computing units, to implement CNN and RNN. By leveraging the static and dynamic responsiveness of sensors, this study implements a spatiotemporal dual-stream network algorithm on a unified hardware platform, successfully overcoming accuracy loss induced by device non-idealities and ultimately achieving near-software-level accuracy (97.8%) on the MNIST dataset. Additionally, the BEOL-compatible ISMC not only exhibits potential for 3D stacking but also realizes hardware-algorithm co-optimization, considering trade-offs between power and accuracy, highlighting its strong potential for reconfigurable, high-performance machine vision.**

Keywords—in-sensor-memory computing, dual-stream algorithm, technology-architecture-algorithm (TAA) design, spatiotemporal machine vision

I. INTRODUCTION

With the rapid development of machine vision technology, particularly in the context of exponentially growing data volumes in the artificial intelligence (AI) era, the development of image sensing systems featuring low power consumption, low latency, and low circuit complexity has become increasingly crucial [1], [2]. However, conventional von Neumann architecture-based visual computing systems, employing a separated processing paradigm of sensing, storage, and computing modules, not only generate substantial redundant data but also introduce significant processing delays [3]. In-sensor computing offers a more energy- and area-efficient computing paradigm integrating data generation, collection, and computation within sensory devices [4], but it faces challenges in deploying on a single hardware platform with versatile functionality. It requires a high consistency of sensory devices, processing systems and computing algorithms, particularly for multi-dimensional data in machine vision, limited by restricted sensing mode, non-linearity, and accuracy degradation [5], [6], [7].

Thus, a cross-layer synergistic design is essential in exploring machine vision encoding and processing tasks, with key metrics (e.g., accuracy degradation, power consumption) harmonized. As shown in Fig. 1, we present an IGO-based phototransistor array with dual-modal photo-response characteristics, enabling spatiotemporal feature extraction from MNIST. These specific characteristics enable a dual-stream convolutional and reservoir computing neural network (C-RNN) with feature fusion. Furthermore, a long-retention IGZO-based 2T0C-DRAM array is implemented to perform network classifier functions with high energy efficiency. This in-sensor-memory computing (ISMC) architecture simplifies the network design, minimizes hardware overhead, and achieves low-power operation with near-software-level accuracy. Consequently, it exhibits an effective sensory capability in recognizing both static and event-based images.

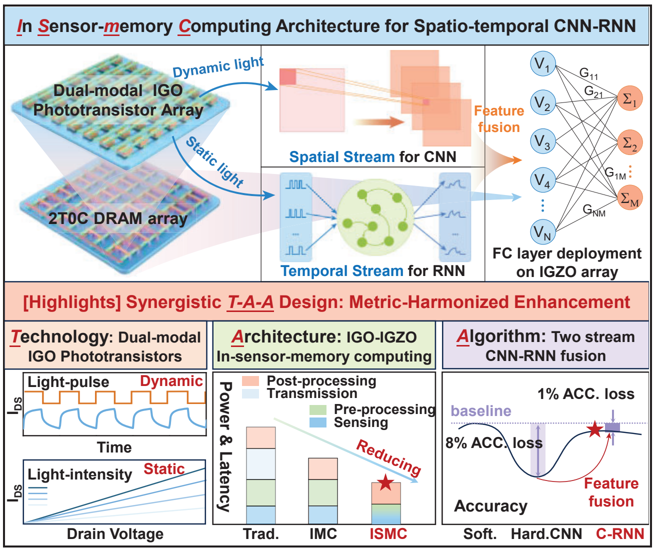

Fig. 1. Neuromorphic vision in-sensor-memory computing for spatiotemporal ANN with synergistic T-A-A design.

II. THE FABRICATION AND BASIC PHOTOTRANSISTOR PROPERTIES OF IGO TFT

A. Device Architecture and Electrical Characterization

A list of fabrication process is illustrated in Fig. 2 (a). P^{++} Si was prepared as substrates. 20 nm Mo was sputtered as gate, patterned by dry etch. 30 nm AlO$_X$ was grown by ALD as the gate dielectric at 300 °C. Before sputtering 10 nm IGO layer at RT, a layer of 5nm SiO$_X$ was deposited at 100 °C. 30nm Ti and 30nm Au were deposited by electron beam evaporation as S/D. 30nm AlO$_X$ was formed by ALD as the passivation at 100 °C. L and W were 30 and 470 µm, respectively. Fig. 2(b) shows the schematic device structure of an a-IGO transistor and the fabrication process flow of the devices. Fig. 2(c)

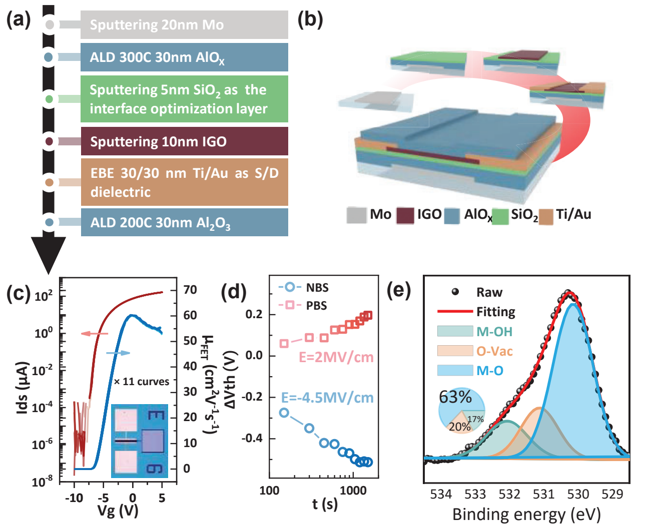

Fig. 2. (a) Fabrication process flow of a-IGO TFTs. (b) Photographs and schematic illustrations of Fabrication process. (c) The transfer characteristics of 11 a-IGO TFTs. (d) Extracted ΔVth of PBS and NBS results for a-IGO devices at 300 K, demonstrating robust bias-temperature instability (BTI) performance. (e) XPS results of a-IGO film.

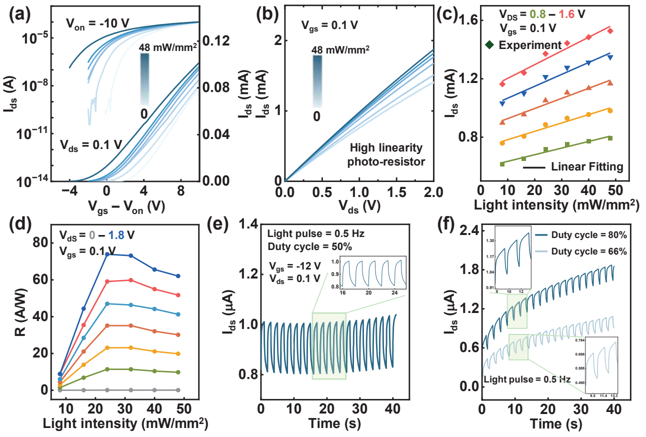

Fig. 3. The light intensity dependent (a) transfer curves and (b) output performance of a-IGO TFTs. (c) Variation of I_{ds} with light intensity in a-IGO TFT at different operating function states. (d) Extracted photoresponsivity R as a function of V_{ds} and light intensity. Dynamic photocurrent in sub-threshold region: (e) 50% duty cycle, (f) 66% and 80% duty cycles.

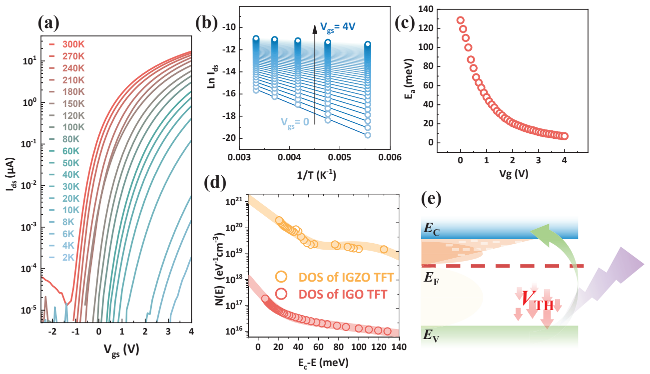

Fig. 4. (a) Temperature-dependent transfer curves of a-IGO TFT from 300K to 2K. (b) The I_{ds} with variation of temperature under different V_{gs} in a-IGO TFT; solid lines are fitted by Arrhenius relation for Ea extraction. (c) The extracted Ea under different V_{gs}. (d) Sub-gap DOS at dielectric/channel interface in IGZO and IGO TFT. Fitting curves were showed in solid lines. The band tail DOS in IGO TFT is significantly lower than that in IGZO TFT. (e) Illustration sketch of the illumination induced instability under visible blue-violet light (405 nm, ~3 eV) illumination conditions.

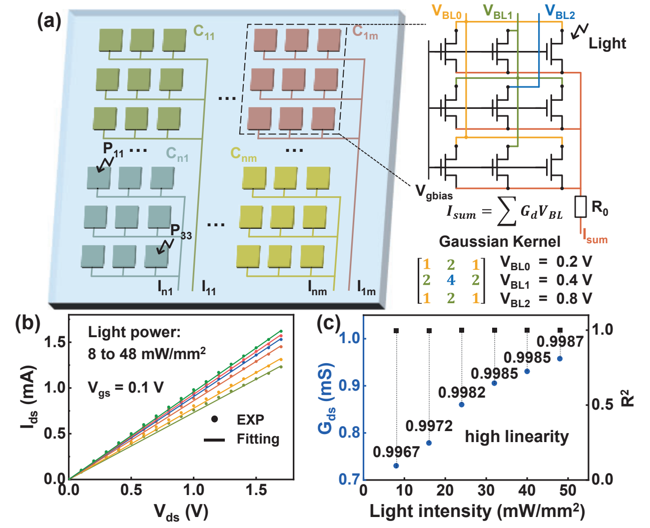

Fig. 5. (a) IGO phototransistors array perform convolution operation with Gaussian kernel. (b) Output characteristics for various blue-violet light powers. (c) Fitting G_{ds} results and R^2 to different light intensity.

presents the transfer characteristics of 11 a-IGO TFTs demonstrates good uniformity with μ_{FET} up to 60 cm^{2}V^{-1}s^{-1}. The PBS and NBS under a stress electric field of 2 and -4.5 MV/cm over 1500 s at room temperature was conducted. The ΔV_{th} was extracted in Fig. 2(d) demonstrates good stability. X-ray photoelectron spectroscopy (XPS) analysis of a-IGO film was presented in Fig. 2(e), indicating the ratios of M-OH, O-Vac and M-O were 17%, 20% and 63%, respectively. Additionally, the real atom ratio of In:Ga is 25:1.

B. *Device Photoresponse and Generation Mechanism*

Fig. 3(a) presents the phototransistor's transfer curves under 405nm blue-violet laser illumination (8-48 mW/mm^2), exhibiting a pronounced leftward V_{th} shift with increasing power due to photogenerated carriers. The observed red and green light responses exhibit similar trends to those reported in Ref. [1]. Additionally, Fig. 3(b) highlights the progressive enhancement of the photocurrent under various illumination intensities within the 8 to 48 mW/mm^2 range. This enhancement under blue-violet laser was attributed to the abundant generation of photogenerated carriers during extensive light exposure. The trends of photocurrent and light intensity under various operating conditions are counted in Fig. 3(c), which shows outstanding linearity of photocurrent in the presence of varying light intensity. Fig. 3(d) shows that the IGO phototransistor exhibits a high photoresponsivity (7.3×10^4 mA/W). Fig. 3(e) and (f) show the nonlinear dynamic photocurrent in the sub-threshold region at 50%, 66%, and 80% duty cycles, respectively, highlighting the effect of varying duty cycle ratios on device current.

The temperature dependence experiment from 300K to 2K reveals the large number of photogenerated carriers and no severe subthreshold decay the transport property in Fig. 4(a). The temperature dependent I_{ds} under different V_{gs} is plotted in Fig. 4(b). Fig. 4(c) shows the activation energy (E_a) in IGO is obtained by fitting the I_{ds} in terms of Arrhenius relationship $I_{ds}\sim Aexp(-E_a/k_BT)$. where A is the preexponential factor. The mobility edge framework was widely used to quantify the density of states (DOS) of channel. Here, E_C-E_F (E_a) is the activation energy for the hopping of electrons. The extracted DOS of band tail states N(E) below the E_C is plotted in Fig. 4(d). The fitting function was:

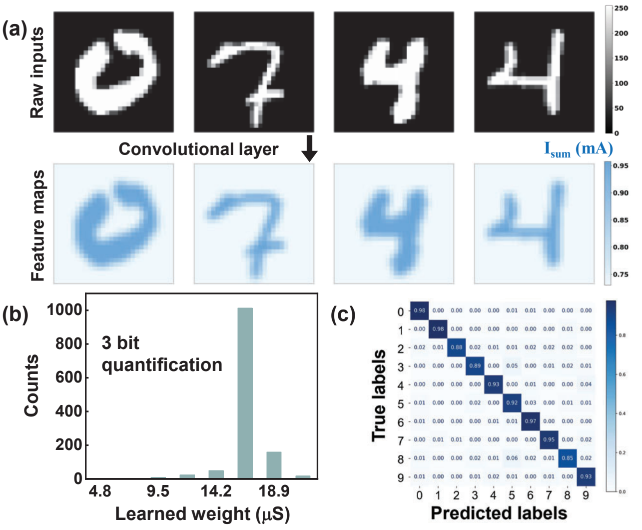

Fig. 6. (a) Edge features extracted from four optical MNIST images illuminated by gaussian convolutional kernel. (b) FC layer weights are determined during offline learning. (c) Confusion matrix.

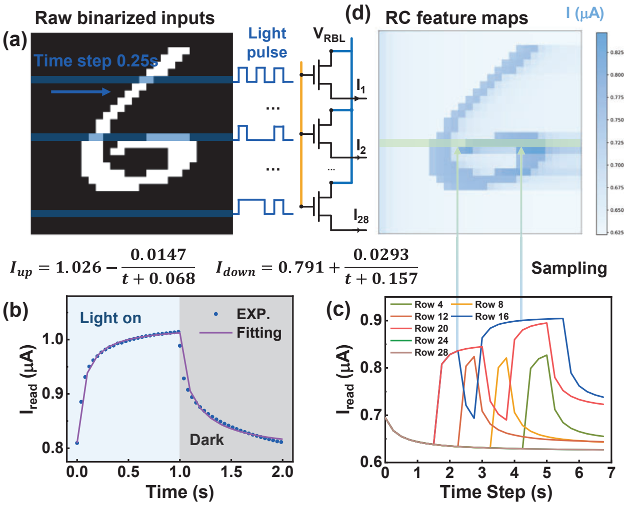

Fig. 7. (a) Schematic illustration of the real-time scanning process of a dynamic image. (b) Dynamic light response and the fitting result. (c) IGO phototransistors array reservoir states for 28 time steps. (d) RC feature maps extracted by sampling transistor current.

$$N(E) = \frac{C_i}{qa}\left(\frac{dE_a}{dV_{gs}}\right)^{-1} \qquad (1)$$

$$N(E) = N_0 \exp\left(\frac{E_C - E_F}{E_U}\right) + N_{00} exp\left(\frac{(E_C - E_F - E_d)^2}{\sigma^2}\right) \qquad (2)$$

where N_0 and N_{00} are the pre-exponential factors, E_U is the Urbach energy, E_d is the donor energy, σ is standard deviation in Gaussian function. Compared to the results of IGZO, a-IGO has less both tail states and Gaussian donor density of states, even by two orders of magnitude. Illustration sketch of the illumination induced instability of a-IGO was in Fig. 4(e).

III. OPTICAL C-RNN IMPLEMENTATION

A. Spatiotemporal Feature Fusion Algorithm

Convolutional layer is deployed in the IGO phototransistors array. We configured a 3×3 IGO array capable of implementing diverse convolutional kernels through precise drain voltage modulation. To validate the feature extraction capability, a Gaussian kernel was specifically constructed for image convolution analysis, as shown in Fig. 5(a). The transistor drain current exhibits a good linearity with respect to the voltage under different light intensities as shown in Fig. 5(b) and (c). The device's superior linear characteristics provide critical assurance for the quantization of convolutional kernel weight multiplications. Fig. 6(a) presents an example of edge detection in four images using the phototransistors array-based 3×3 Gaussian kernel. The learned weights were encoded as 3-bit states for the 2T0C DRAM array-based fully connected (FC) layer, as shown in Fig. 6(b). Fig. 6(c) presents the confusion matrix of the single-layer CNN, demonstrating a recognition accuracy of 93.2%.

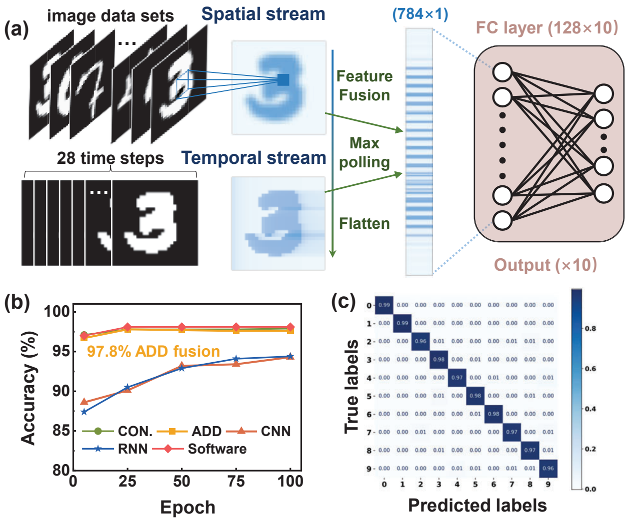

Fig. 8. (a) Diagram illustrating the fusion of spatial and temporal stream features. (b) Comparison of accuracy across five different algorithms. (c) Confusion matrix in add feature fusion approach.

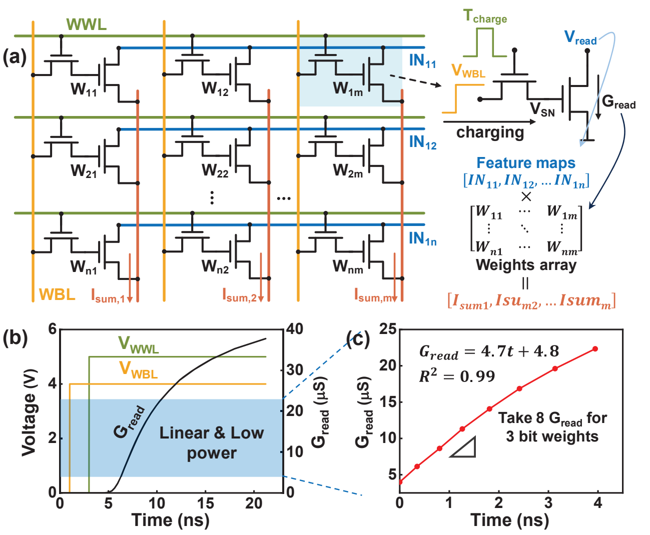

Fig. 9. (a) 2T0C IGZO DRAM array to implement the FC layer analog computing. (b) Weight writing operation through charging read transistor conductance. (c) Linear region for 3-bit quantification.

The reservoir computing (RC) layer is implemented using the dynamic light response of IGO transistors. Sampling the nonlinear current response of the transistor allows for the extraction of temporal features from the input binarized images as shown in Fig. 7(a). The photocurrent fitting of the transistor is shown in Fig. 7(b). The nonlinear photocurrent response of the IGO optical transistor was characterized with a scanning temporal resolution of 0.25s over a 7s duration, as

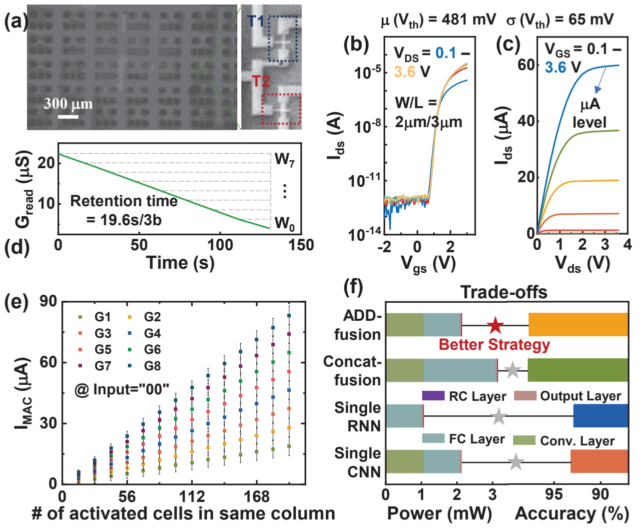

Fig. 10. (a) SEM images of the DRAM array. (b) Transfer (c) output characteristics and (d) retention time of IGZO TFT. (e) MAC current distribution across different transconductance (G1~8). (f) Trade-offs between power and accuracy for four algorithms.

TABLE I. BENCHMARK OF THE PROPOSED IGO BASED SENSOR ARRAY COMPARED WITH OTHER PROCESSING IN SENSOR METHODS

	IEDM 2023 [2]	**ESSERC 2024 [5]**	**Nature 2020 [6]**	**IEDM 2022 [9]**	**This work**
Sensor Type	In$_2$O$_3$ PT - NbO$_x$ TS	CMOS 180nm	WSe$_2$ Diode	FE-HZO Cap	**IGO Transistor**
Responsivity (mA/W)	10^4	NONE	60	NONE	**$7.3*10^4$**
Task	Fashion-MNIST	MNIST	MNIST	Motion Pattern Detection	**MNIST**
Accuracy	82%	95.1%	92.8%	99%	**97.8%**
Operation	SNN	4-layer CNN	Classifier and autoencoder	FC layer classifier	**CNN + RNN + classifier (2T0C)**
Reconfigurable NN	No	No	No	No	**Yes**

shown in Fig. 7(c). Using identical sampling resolution, the acquired RC feature map is presented in Fig. 7(d). A single-layer recurrent neural network leveraging these characteristics achieved 92.9% recognition accuracy.

B. *Dual-stream Neural Network Implementation*

The IGO array can achieve static and dynamic feature extraction of images, but the recognition accuracy of single convolutional or reservoir computing layer cannot reach software-level recognition accuracy due to the non-ideal characteristics of the devices. By fusing the temporal and spatial features of the MNIST dataset images, the dual-stream neural network is trained and inferred through the feature fusion, max pooling, flatten, and fully connected layer, as shown in Fig. 8(a). The accuracy of the C-RNN was improved by 4% over the single neural network, achieving software-level performance, as shown in Fig. 8(b) and (c).

Fig. 9(a) shows the 2T0C DRAM in-memory computation array used to implement the fully connected layer for classifier tasks. The weights of the fully connected layer are written by charging the conductance of the read transistor. The conductance value exhibits good linearity with the pulse width of the write word line, and eight analog values are selected for 3-bit quantization of the weights, as shown in Fig. 9(b) and (c). Fig. 10(a) shows SEM images of the DRAM array, and the process refers to [8]. The transfer, output characteristics, and retention time of the IGZO TFT are shown in Fig. 10(b), (c) and (d). Fig.10(e) shows the I_{MAC} distribution as a function of the number of activated 2T0C cells in same column. Fig. 10(f) shows the trade-offs between power and accuracy for four algorithms. The dual-stream NN achieves higher accuracy; however, different fusion strategies result in varying power consumption. Add fusion enables lower power while maintaining a comparable level of accuracy, making it a better fusion strategy.

IV. CONCLUSION

This work demonstrates an IGO-based in-sensor computing architecture that integrates sensing and encoding on a unified platform, enabling spatiotemporal feature extraction for diverse neural networks. Our convolutional-reservoir hybrid algorithm compensates for device non-idealities, boosting MNIST accuracy by 4% to 97.8% (Table I) with negligible power overhead. The integration of IGO sensing array with IGZO-2T0C-based FC network classifiers enables low-complexity image ISMC systems. Our approach exhibits ultra-high energy efficiency potential for visual edge computing applications.

ACKNOWLEDGMENT

This work was supported in part by National Key Research and Development Program of China under Grant 2023YFB4402400; National Natural Science Foundation of China under Grants 92464201, U2341218, 62274178, 62322412.

REFERENCES

[1] Dang, B. et al., "Reconfigurable in-sensor processing based on a multi-phototransistor-one-memristor array," *Nature Electron.*, vol. 7, no. 11, pp. 991-1003, Nov. 2024.

[2] R. Wang et al., "1-Phototransistor-1-Threshold Switching Optoelectronic Neuron for In-Sensor Compression via Spiking Neuron Network," in *IEDM Tech. Dig.*, pp. 31.1.1-33.1.4, 2023.

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

[4] F. Zhou and Y. Chai, "Near-sensor and in-sensor computing," *Nature Electron.*, vol. 3, no. 11, pp. 225-232, Nov. 2020.

[5] M. Nazhamaiti et al., "Selfputing: A 0.57 μW @ 15 fps Vision Chip with Self-powered In-Pixel Computing and In-Memory Computing for Visual Perception on the Edge," 2024 *IEEE European Solid-State Electronics Research Conference (ESSERC)*, 2024, pp. 585-588.

[6] L. Mennel et al., "Ultrafast machine vision with 2D material neural network image sensors," *Nature*, vol. 579, no. 7797, pp. 62-66, Mar. 2020.

[7] X. Wang et al., "6.9 A 0.35V 0.367TOPS/W Image Sensor with 3-Layer Optical-Electronic Hybrid Convolutional Neural Network," in *Proc. IEEE Int. Solid-State Circuits Conf. (ISSCC)*, 2024, pp. 116-118.

[8] G. Yan et al., "First Demonstration of True 4-bit Memory with Record High Multibit Retention $>10^3$s and Read Window $>10^5$ by Hydrogen Self-Adaptive-Doping for IGZO DRAM Arrays," in *IEDM Tech. Dig.*, pp. 6.7.1-6.7.4, 2023.

[9] Z. Fu et al., "Novel Energy-efficient Hafnia-based Ferroelectric Processing-in-Sensor with in-situ Motion Detection and Four-quarter Mutipilcation, " in *IEDM Tech. Dig.*, pp. 24.5.1-24.5.4, 2022.