Nanoscale Horizons

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

View Article Online
View Journal

Cite this: DOI: 10.1039/d5nh00720h

Received 29th October 2025,
Accepted 6th January 2026

DOI: 10.1039/d5nh00720h

rsc.li/nanoscale-horizons

An IGZO phototransistor-based ternary inverter integrating optical sensing and weight quantization in ternary neural networks for color image recognition

Wun-Yun Lin,[a] Yong-Yi Huang,[a] Yu-Chieh Chen,[a] Chen-Gang Jang,[b] Li-Chung Shih [a] and Jen-Sue Chen *[ab]

As deep neural networks (DNNs) continue to advance computer vision, natural language processing, and medical diagnostics, their reliance on 32-bit full-precision weights imposes substantial model size and computational burdens that hinder deployment at the edge; to improve efficiency, we adopt ternary neural networks (TNNs). Here, we present a ternary circuit composed of two parallel indium–gallium–zinc-oxide (IGZO) thin-film phototransistors (TFPTs) and a resistor, exhibiting three stable, discrete current states 'OFF', 'Intermediate', and 'ON' that map to the ternary weight set {−1, 0, 1}; we further realize a compact ternary inverter using only two IGZO TFPTs and two resistors, avoiding complex binary CMOS logic. The processing path begins with optical sensing, wherein the incident light power densities and wavelength determine discrete voltage outputs; during preprocessing, these voltages discretize pixel values (0–255) into multiple intervals that are supplied to the TNN for image recognition. Leveraging this integrated sensing, preprocessing and inference hardware module, we achieve >90% accuracy on CIFAR-10, thereby validating device-level data discretization and transformation and charting a path toward integrated neuromorphic vision systems.

New concepts

This work demonstrates that low-cost and scalable IGZO thin-film transistors, with turn-on voltages engineered by sputtering process modulation, can be assembled into a simple parallel circuit of IGZO devices and resistors to realize both ternary *I–V* characteristics and a ternary inverter under dark conditions. Moreover, owing to the intrinsic photoresponsivity of IGZO, RGB illumination directly modulates the voltage transfer behavior, enabling pixel-to-voltage mapping without additional conversion circuits. By integrating ternary weight quantization, optoelectronic sensing, and preprocessing within ternary neural networks, the proposed circuits achieve over 90% accuracy on CIFAR-10, highlighting a practical route toward scalable neuromorphic hardware.

Introduction

Conventional hardware implementations of neural networks predominantly rely on Complementary Metal-Oxide-Semiconductor (CMOS) technology,[1,2] which is inherently binary in nature. Such binary logic behavior restricts devices to two discrete states, '0' and '1', limiting the ability to efficiently realize multi-level states and full-precision weights that are critical for advanced neural computations. Implementing these functions on binary CMOS circuits often necessitates additional peripheral components such as comparators, multiplexers, or memory units.[3–5] This substantially increases circuit complexity, power consumption, and footprint. Moreover, the prevalent use of 32-bit full-precision weights in DNNs results in models with hundreds of millions of parameters,[6,7] demanding extensive computational resources in pursuit of high predictive accuracy. These factors hinder the deployment of neural networks on resource-constrained devices. To address this problem, TNNs offer a compelling alternative by quantizing weights into three discrete levels,[8,9] represented as {−1, 0, 1}. By adopting ternary quantization, the data representation is nearly compressed by a factor of sixteen compared to full precision,[10] while maintaining competitive accuracy. Despite these advantages, direct hardware realization of ternary weight quantization remains challenging. Many existing solutions rely on binary CMOS circuits augmented with complex control logic to emulate ternary states.[11–13]

To address these challenges, we propose a compact circuit based on parallel IGZO TFPTs with engineered turn-on voltages optimized *via* sputtering process control, combined with a resistor. This configuration produces a stable intermediate current state (∼100 nA) within an intermediate voltage range (−3 to 0 V). We directly encode the three stable discrete current states observed in the *I–V* characteristics as ternary weights

[a] *Department of Materials Science and Engineering, National Cheng Kung University, Tainan, 70101, Taiwan. E-mail: jenschen@ncku.edu.tw*
[b] *Program on Semiconductor Packaging and Testing, Academy of Innovative Semiconductor and Sustainable Manufacturing, National Cheng Kung University, Tainan, 70101, Taiwan*

{−1, 0, 1}, serving as hardware elements for TNNs.[10,14–16] Furthermore, by utilizing the parallel device configuration with two resistors, we demonstrate a ternary inverter that implements a multi-valued logic system[17–19] with three distinct output voltages corresponding to logic states '1', '0', and '−1', typically requiring more complex binary CMOS hardware. This intrinsic ternary behavior demonstrates its potential as hardware components for TNN activation functions and classifiers. To further situate the proposed approach, a concise comparison with representative multi-level device platforms is provided in Table S1. This comparison highlights the advantage of integrating photosensing and ternary behavior within a parallel IGZO device configuration.

Moreover, the ternary inverter exhibits optoelectronic modulation characteristics. Under RGB laser illumination, its output voltage transfer characteristics are tunable by incident light power density and wavelength. This functionality enables direct mapping of pixel values into discrete voltage signals, realizing sensing and encoding[20,21] without the need for complex external conversion. The photoresponsive voltage outputs can be applied in CIFAR-10 classification tasks, further demonstrating the potential for integrating sensing and computation within a IGZO parallel device.

In summary, we demonstrate a compact circuit that integrates optoelectronic sensing with inherent ternary logic behavior. The IGZO parallel device, with its stable three-state current levels, serves not only as a weight quantization element but also as a foundation for designing a photosensitive ternary inverter as a multi-valued logic element for sensing. These multifunctional capabilities enable signal processing at the circuit-level, alleviating the complexity of subsequent digital computations. Such advantages establish the device as a foundational component for practical vision-driven neuromorphic hardware systems, including edge computing and real-time image analysis. Our approach paves the way toward energy-efficient, scalable hardware architectures that unify sensing and computing, advancing the development of next-generation artificial intelligence hardware.

Results and discussion

We present a parallel IGZO TFPT structure, fabricated through strategically controlled sputtering parameters to achieve precise deposition of IGZO films, enabling channel layers with functionally distinct physical characteristics. This structural configuration enables differential electrical performance. Fig. S1 illustrates the schematic design of the parallel IGZO TFPTs and presents an optical microscopy (OM) image of the fabricated device. The parallel IGZO TFPTs share a common source node and a common gate electrode, and the gate voltage (V_G) is designated to be the input voltage (V_{in}). When combined within a properly designed circuit, the system functions as a weight quantization unit, delivering three discrete and stable output current states ('OFF', 'Intermediate', and 'ON'),[10,14–16] as illustrated in Fig. 1(a). Such behavior supports the hardware realization of ternary quantization weights for neural networks.

Beyond its electrical operation, the device also exhibits photoresponsive characteristics relevant to photonic signal preprocessing.[22,23] Under tunable RGB laser illumination, as illustrated in Fig. 1(b), photo-induced carriers modulate the effective gate potential, producing a wavelength- and light power density-dependent shift in the first transition voltage. The corresponding voltage transfer curves reveal distinct transitions associated with specific color channels and illumination densities, enabling color encoding of optical stimuli into electrical signals.

This dual-mode operation may facilitate compact implementations of sensing and computing units for circuit-level encoding and inference in hardware neural networks. In Fig. 1(c) and (d), cross-sectional TEM images confirm well-defined IGZO/SiO_2/Si interfaces for devices with IGZO sputtered at 50 W and 70 W, respectively. Elemental maps in Fig. 1(e) and (f) further validate the uniform distribution of O, Si, In, Ga, and Zn, indicating consistent film quality across both device configurations.

To investigate the impact of sputtering power on the electrical properties of amorphous oxide semiconductors, IGZO TFPTs were fabricated using two different sputtering powers, corresponding to different turn-on voltages. As shown in Fig. 2(a), the devices adopt a bottom-gate structure with a 100 nm thermally grown SiO_2 layer serving as the gate dielectric. The characteristics of the two TFPTs are presented in Fig. 2(b). With increasing sputtering power, the turn-on voltage exhibits a negative shift of approximately −5 V. The observed negative shift in turn-on voltage by increasing the sputtering power may be attributed to the increased defect density due to an elevated resputtering effect.[24] To systematically strengthen our conclusion regarding the resputtering effect, we also deposited IGZO devices at sputtering powers of 30 W and 100 W. The corresponding electrical properties are provided in Fig. S2. The TFPTs with IGZO sputtered at 50 W and 70 W are labelled as IGZO–$V_{on,o}$ and IGZO–$V_{on,shift}$, respectively. The corresponding electrical parameters, including threshold voltage (V_{th}) and subthreshold swing (S.S.), are summarized in Fig. S3. All measurements were conducted under identical bias conditions using continuous gate voltage sweeps.

In addition, the UV-vis transmittance spectra of the IGZO films deposited at different sputtering powers are shown in Fig. S4. The optical bandgap (E_g) was estimated using the Tauc plot method by fitting the linear region of the $(\alpha h\nu)^2$ *versus* $h\nu$ plot, where α is the absorption coefficient and $h\nu$ is the photon energy. The extracted E_g values were approximately 3.6 eV for the IGZO–$V_{on,o}$ IGZO film and 3.7 eV for the IGZO–$V_{on,shift}$ film. Ultraviolet photoelectron spectroscopy (UPS) was further employed to determine the work function (Φ_{IGZO}), which was found to be 3.92 eV and 3.83 eV for the IGZO–$V_{on,o}$ and IGZO–$V_{on,shift}$ layers, respectively, as shown in Fig. S5. The energy difference between the Fermi level and the valence band maximum ($E_F - E_V$) was calculated to be 3.19 eV for the IGZO–$V_{on,o}$ film and 3.25 eV for the 35 nm film. Using the relationship $\Phi_{IGZO} - [E_{g,IGZO} - (E_{F,IGZO} - E_{V,IGZO})]$, the conduction band minimum (E_C) relative to the vacuum level was determined to be 3.51 eV for the IGZO–$V_{on,o}$ film and 3.38 eV for the IGZO–$V_{on,shift}$ film. Based on these values, the equilibrium band diagrams of the IGZO–$V_{on,o}$ and $V_{on,shift}$ layers were constructed, as illustrated in Fig. 2(c) and (d).

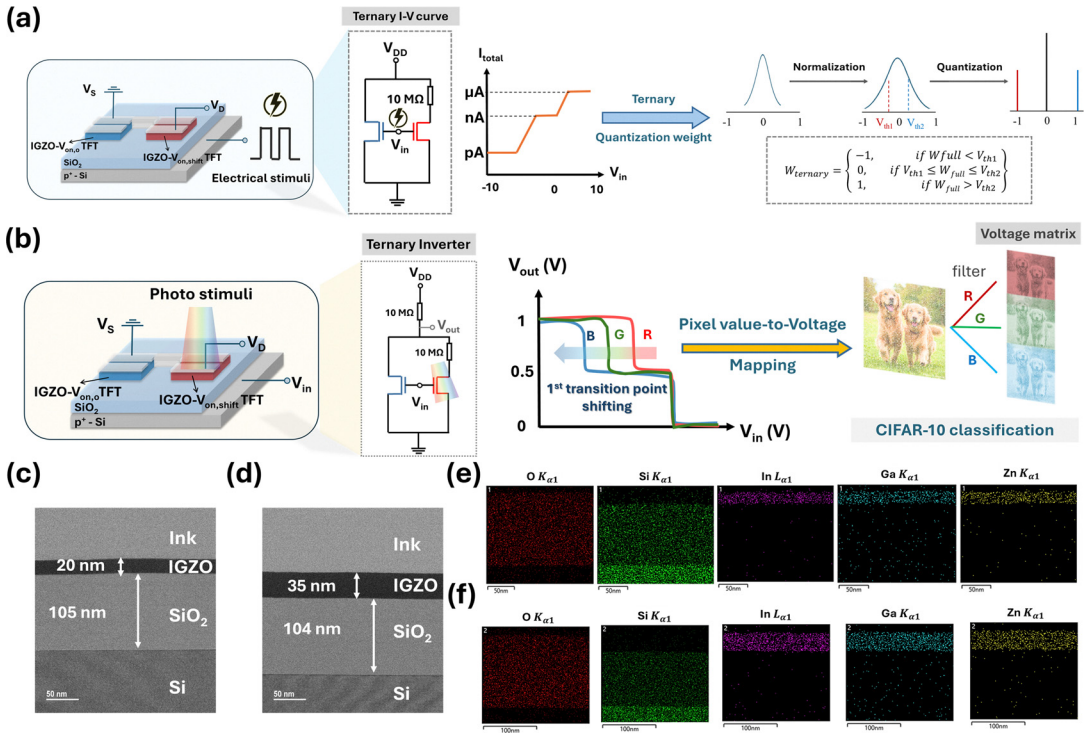

Fig. 1 Illustrations and material analyses of an IGZO-based parallel device for integrated electrical and optical processing. (a) Schematic and operation of the ternary circuit under electrical stimuli. The device structure consists of two IGZO thin-film phototransistors (TFPTs) with different turn-on voltages (IGZO–$V_{on,o}$ and IGZO–$V_{on,shift}$) and exhibits a characteristic ternary I_D–V_G transfer curve, enabling the mapping of output current (I_{total}) into weight values {−1, 0, 1}. (b) The device operation under optical stimuli demonstrating wavelength-dependent voltage transfer characteristics, enabling a pixel value-to-voltage mapping for image classification tasks. Cross-sectional TEM images with (c) IGZO–$V_{on,o}$ and (d) IGZO–$V_{on,shift}$ active layers, showing clear IGZO/SiO_2/p^+–Si stacking. EDS elemental maps correspond to (e) IGZO–$V_{on,o}$ and (f) IGZO–$V_{on,shift}$ active layers, confirming uniform distribution of O, Si, In, Ga, and Zn in the IGZO active layers.

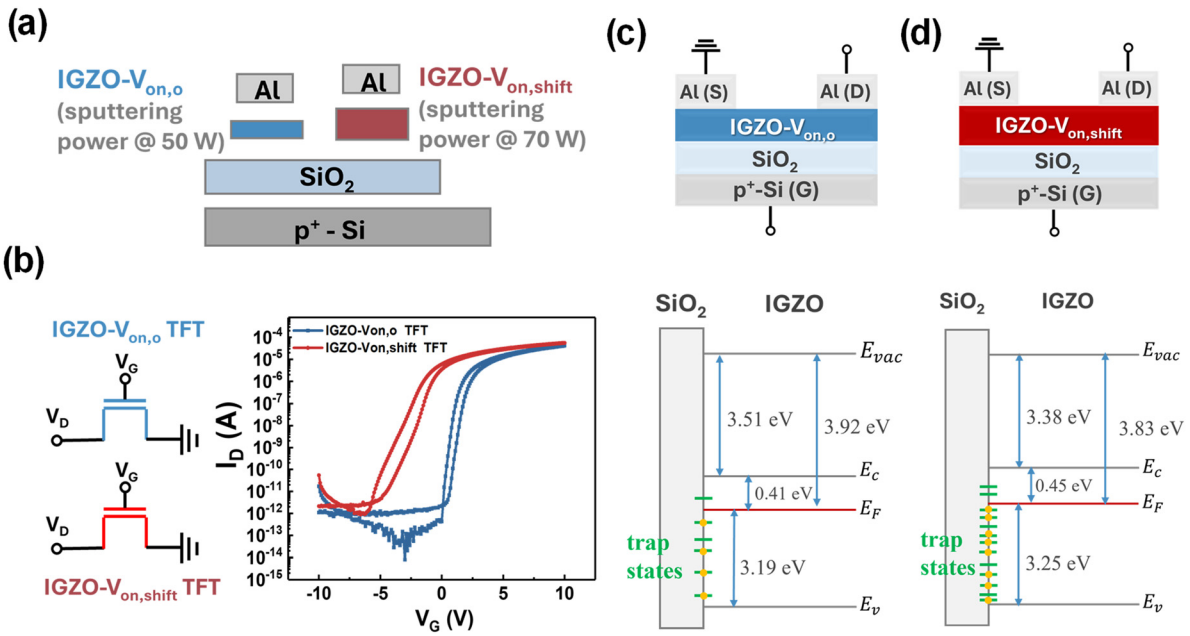

Fig. 2 Electrical performance and compositional properties of IGZO TFPTs and the corresponding energy band diagrams. (a) Schematic cross-section of IGZO TFPTs with IGZO–$V_{on,o}$ and $V_{on,shift}$ active layers, fabricated on SiO_2/p^+–Si substrates with Al electrodes. (b) Transfer characteristics (I_D–V_G) of the IGZO–$V_{on,o}$ (blue) and $V_{on,shift}$ (red) IGZO TFPTs measured at V_D = 1 V, using a bidirectional gate voltage sweep from −10 V to +10 V and back to −10 V. Energy band diagrams of (c) IGZO–$V_{on,o}$ and (d) IGZO–$V_{on,shift}$ interfaces under zero gate bias conditions, illustrating a higher density of trap states at the interface in the IGZO–$V_{on,shift}$ film.

Fig. 3(a) shows the equivalent circuit of a parallel IGZO TFPT structure used to demonstrate three stable current states. The IGZO–$V_{on,shift}$ TFPT begins to turn on at approximately V_G = −5 V and is connected in series with a 10 MΩ resistor, which limits the current to around 100 nA when the gate voltage is between −3 V and 0 V, forming an intermediate conduction state. As V_G increases beyond 0 V, the 20 nm IGZO TFPT on the left branch also turns on. Since this path has no current-limiting resistor, the current preferentially flows through it, resulting in a clear transition to the ON state. The I_D–V_G curve shows three distinct current levels: 'OFF', 'Intermediate' and 'ON' states. This behavior highlights the potential of the circuit for ternary logic systems and multi-level neuromorphic applications. At a supply voltage of V_{DD} = 1 V, the three current states correspond to estimated static power consumptions of approximately 20 pW, 95 nW, and 70 μW, respectively. To examine the cycle-to-cycle (C2C) and device-to-device variations (D2D) of the parallel IGZO TFPTs, the I_D–V_G curves and gate current (I_G) were measured in the dark. As shown in Fig. S6 and S7, the results indicated that the values of I_G for the parallel IGZO TFPTs were significantly smaller than the drain current (I_D), revealing that the device's leakage current could be considered negligible. Detailed electrical parameters extracted from the C2C and D2D tests are summarized in Tables S2 and S3, respectively, providing further insights into the device stability and uniformity. Fig. 3(b) shows the cumulative distribution function (CDF) of the drain current (I_D) measured under three gate voltage conditions (−10 V, −1 V, and +10 V), corresponding to the 'OFF', 'Intermediate' and 'ON' states, respectively. After 100 repeated measurements, the data demonstrate robust state separation

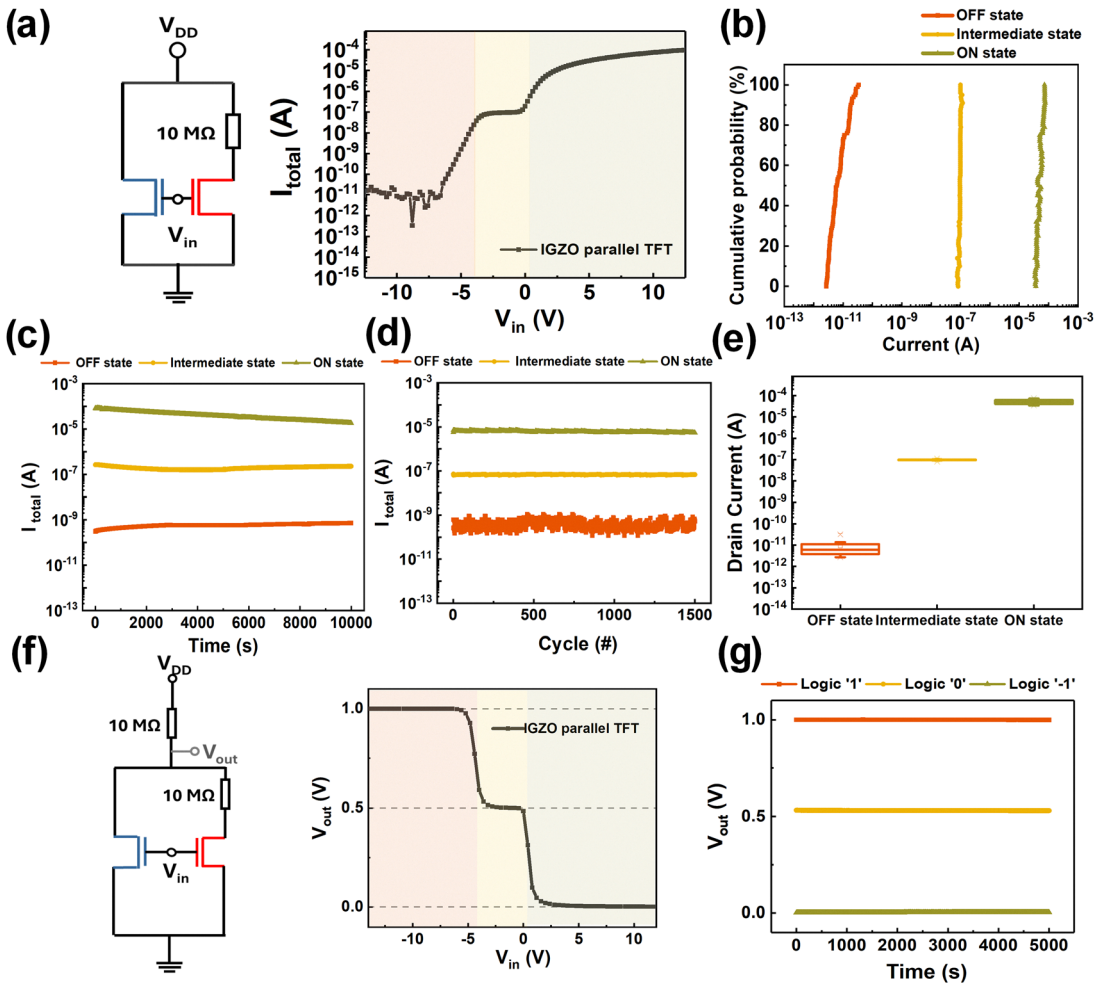

Fig. 3 Characterization of the ternary current state and inverter circuits. (a) Ternary I_D–V_G transfer curve with circuit diagram, showing the measured total drain current (I_{total}) as a function of gate voltage (V_G) at a drain voltage (V_D) of 1 V. (b) Cumulative probability distribution of drain current levels at −10 V, −1 V and 10 V gate voltages, corresponding to the OFF/Intermediate/ON current states, respectively, measured over 100 repeated measurements during the V_G sweep from −10 V to 10 V. (c) Retention characteristics of each state measured for 10 000 seconds. (d) Endurance test of the ternary states over 1500 cycles, using a repeated pulse sequence of V_G = −5 V ('OFF'), −0.5 V ('Intermediate'), and 5 V ('ON'), each with a pulse width of 100 ms at a drain voltage (V_D) of 1 V. (e) Distribution of the drain currents in each state, where the box shows the data between 25 and 75%. The mean value and standard deviation of the data are shown as hollow squares and error bars. (f) The VTC curve of a ternary inverter along with its circuit diagram, at an operating voltage (V_{DD}) of 1 V. (g) Retention characteristics of the three distinct logic states: '1', '0' and '−1', corresponding to gate voltages of −10 V, −1 V, and 10 V, respectively.

with minimal overlap, indicating stable and reliable reproducibility under repeated measurements. To evaluate temporal retention, Fig. 3(c) plots the drain current over 10 000 seconds under the three gate voltage biases. The Intermediate state, which represents a crucial quantization level between 'ON' and 'OFF', remains highly stable with no noticeable degradation, confirming the retention capability of the device for long-term triple state representation. Fig. 3(d) further investigates the endurance characteristics by applying 1500 cycles of alternating gate voltage pulses (V_G = −5 V, −0.5 V, and 5 V, each pulse width = 100 ms) corresponding to the three programmed states. Each pulse was set to a distinct gate voltage associated with 'OFF', 'Intermediate', and 'ON' current states, respectively. The resulting current levels remain highly stable throughout the repeated switching cycles, and the measured drain currents remained stable across all states, confirming reliable multi-level switching behavior. Statistical analysis of the measured drain current distributions is shown in Fig. 3(e), after 100 V_G sweeps from −10 V to 10 V. The extracted mean and standard deviation (μ, σ) for the 'OFF', 'Intermediate', and 'ON' states are (2.0×10^{-11} A, 1.17×10^{-11} A), (9.50×10^{-8} A, 2.92×10^{-9} A), and (7.0×10^{-5} A, 5.83×10^{-6} A), respectively. The small σ values across all states confirm the high precision and reliability of the three individual current levels.

In addition, we performed illumination C2C measurements to evaluate repeatability during optical operation. As shown in Fig. S8, the ternary current states remain clearly separated with minimal overlap across repeated cycles under illumination. Furthermore, repeated electrical and optical stress may induce degradation in oxide TFTs through charge trapping and defect related processes.[25,26] In this work, we performed repeated optical exposure and positive gate bias–stress tests as shown in Fig. S9 and S10, and the observed gradual V_{on} shifts are consistent with prior oxide TFT reports. To further enhance long-term stability, adding interfacial layer, post-deposition annealing, and sputtering-condition optimization to suppress oxygen vacancies can be adopted.[27–29]

Fig. 3(f) shows the circuit diagram of a ternary inverter, which serves as the fundamental unit for ternary logic circuits. This inverter is implemented using the same IGZO parallel TFPT configuration and is enhanced by the inclusion of an additional load resistor. The circuit converts the input gate voltage (V_{in}) into three distinct output voltage levels. As shown in the voltage transfer characteristic (VTC) curve, the output voltages V_{out} = 1 V, 0.5 V, and 0 V correspond to the logic states '1', '0', and '−1' with estimated static power consumptions of 1.33 nW, 50.1 nW, and 98.9 nW at V_{DD} =1 V, respectively.

The underlying operating mechanism of this ternary inverter can be explained by examining its behavior across three input voltage regions. When V_{in} is sufficiently negative ($V_{in} \leq -5$ V), both IGZO TFPTs are in the turn-off state. Under this condition, the load resistor behaves effectively as a wire, allowing V_{out} to rise to the supply voltage level (V_{DD} = 1 V), which corresponds to logic state '1'. As V_{in} increases within the intermediate range (-5 V $< V_{in} <$ 0 V), the IGZO–$V_{on,shift}$ TFPT is in the turn-on state while the other remains with high resistance. This results in a voltage divider formed between the load resistor and the conducting IGZO–$V_{on,shift}$ TFPT, which causes V_{out} to stabilize at 0.5 V, corresponding to logic state '0'. When V_{in} becomes positive ($V_{in} > 0$ V), both TFPTs conduct. This causes V_{out} to drop to 0 V, which corresponds to logic state '−1'.

As demonstrated in Fig. S11, the same device can also be configured as a conventional binary inverter. In comparison to the ternary configuration shown in Fig. S12, the binary operation lacks the ability to express intermediate logic levels. This results in reduced functional density and limited logic expressiveness. In contrast, the ternary implementation leverages the multi-threshold switching characteristics of the IGZO parallel device to provide native support for multi-valued logic.

In the proposed ternary neural network (TNN), weights are quantized to three levels: −1, 0, and 1. This simplifies full-precision weights while keeping their sign and direction. Ternary quantization reduces memory and computation by replacing multiply–accumulate operations with simpler add, skip, and subtract steps. It also speeds up inference and lowers power use by making the model sparse. Fig. 3(g) confirms the temporal robustness of the ternary logic output under a constant voltage bias over 5000 s, underscoring the feasibility of TNN-compatible hardware with quantized weight representation and logic computation within the IGZO-based device.

The three-state logic device presented in this study exhibits photosensitive characteristics, enabling precise tuning of logic transition voltages through optical stimulation. This behavior originates from the intrinsic properties of amorphous oxide thin-film phototransistors,[30,31] and their inherently low dark carrier concentration, which together result in photoconductive responses under visible-light illumination. These properties allow efficient modulation of channel conductivity by incident light, making the devices suitable for visible-light sensing and optoelectronic logic applications. As illustrated in Fig. 4(a), the ternary inverter comprises two parallel IGZO TFPTs and two 10 MΩ resistors connected in series to V_{DD}, with the output voltage (V_{out}) measured at the node between the resistors. A gate voltage (V_{in}) is applied as the input. In the pristine (dark) state, the device exhibits three distinct and stable logic levels, as shown in the VTC curve in Fig. 4(b)–(d).

To investigate the mechanism behind the light-induced shift of logic transitions, we first compare the photoresponses of two types of IGZO TFPTs. Fig. S13 presents the I_D–V_G characteristics of IGZO–$V_{on,o}$ TFPT under RGB light illumination across increasing light power densities. Minimal turn-on voltage shift under red light is observed, indicating a weak photoresponse and serving as a control group. In contrast, Fig. S14 shows that the IGZO–$V_{on,shift}$ TFPT exhibits a significant and systematic negative shift in turn-on voltage under the same illumination conditions. This pronounced response reveals a strong light-induced modulation capability and supports the presence of a dual-enhancement mechanism involving both negative gate bias and photon-generated carrier dynamics.[25,32]

Under dark equilibrium conditions, the abundant defect states trap free carriers, while neutral oxygen vacancies (V_O) located near the valence band do not contribute mobile carriers.

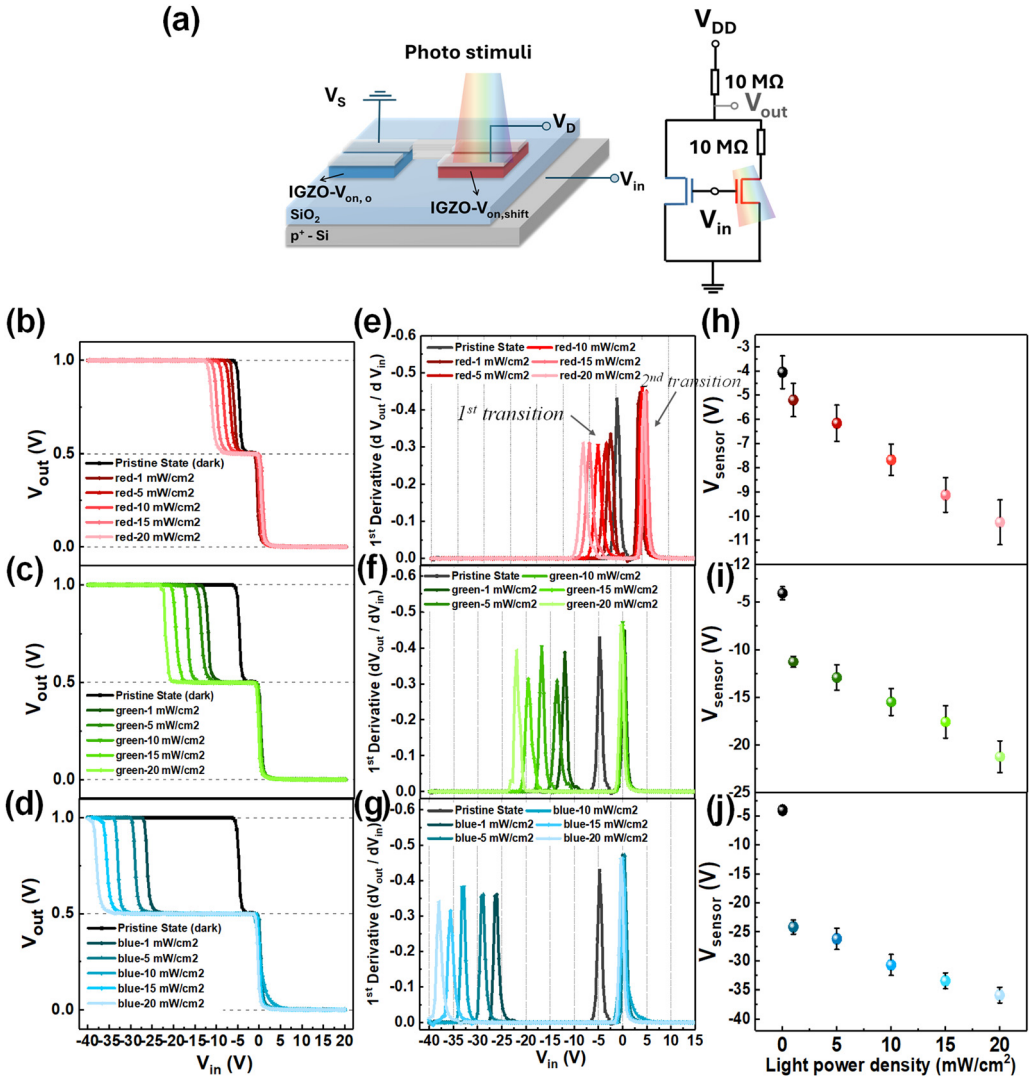

Fig. 4 Photo-modulation of an IGZO-based ternary inverter for RGB feature encoding. (a) Device schematics depicting the IGZO-based ternary inverter under illumination. (b)–(d) Voltage transfer characteristics (VTCs) for varying light power densities of red, green, and blue light (0–20 mW cm^{-2}), demonstrating a clear shift in the inverter's switching behavior. (e)–(g) First derivatives of the VTCs (dV_{out}/dV_{in}) highlight the first transition voltage (V_{sensor}), corresponding to the transition from Logic '1' ($V_{out} \approx 1$ V) to Logic '0' ($V_{out} \approx 0.5$ V). (h)–(j) Relationship between incident light densities (0–20 mW cm^{-2}) and the corresponding V_{sensor} for each color channel, showing monotonic voltage shifts. These voltages are mapped to R, G, and B pixel values as TNN inputs.

Upon visible light illumination, V_O undergoes a dissociation reaction to form ionized V_O^{2+}, releasing two free electrons per vacancy and significantly enhancing channel conductivity. When a negative gate bias is simultaneously applied during illumination, it not only facilitates the release of trapped electrons but also promotes the separation of photogenerated carriers *via* the electric field,[25,32] thereby substantially increasing the carrier concentration and conduction behavior within the channel, which causes a pronounced voltage shift.

To verify the correlation between increased sputtering power and oxygen vacancy formation, we conducted X-ray photoelectron spectroscopy (XPS) analysis of the IGZO films deposited at different powers. The deconvoluted O 1s spectra, shown in Fig. S15, reveal two distinct peaks at binding energies of 530.1 eV and 531.6 eV, corresponding to lattice oxygen and non-lattice oxygen (associated with V_O), respectively.[33,34] The calculated proportion of oxygen vacancies was 14.6% for the film sputtered at lower power (IGZO–$V_{on,o}$) and 18.1% for that sputtered at higher power (IGZO–$V_{on,shift}$), confirming the increased V_O concentration with elevated sputtering power.

This mechanism also explains the selective shift observed in the ternary inverter's first logic transition point. Specifically, the Logic '1' to '0' transition originates from the earlier turn-on behavior of the IGZO–$V_{on,shift}$ TFPT under illumination. Therefore, the shift in the I_D–V_G curve under light directly accounts for why only the first transition point exhibits sensitivity to light power density and wavelength, validating the optoelectronic modulation potential of the device.

Upon illumination with lasers of varying wavelengths and light power densities (1–20 mW cm^{-2}), a systematic shift in the

logic transition voltages is observed. The transition from Logic '1' (high voltage state, V_{out} = 1 V) to Logic '0' (intermediate voltage state, V_{out} = 0.5 V) shifts progressively toward more negative input voltages as the power density increases. As shown in Fig. 4(b), under red illumination (670 nm), the VTC curve exhibits a moderate negative shift with increasing power density. At 20 mW cm^{-2}, the first transition point shifts approximately −6 V from its pristine-state position while the transition remains relatively sharp, maintaining a well-defined switching characteristic. Fig. 4(c) demonstrates that green illumination (520 nm) produces a more pronounced negative shift. The maximum power density (20 mW cm^{-2}) causes the first transition voltage to shift by approximately −17 V from the pristine state, revealing a stronger photo modulation effect compared to red illumination. Fig. 4(d) illustrates that blue illumination (405 nm) induces the most dramatic shift among all tested wavelengths. At 20 mW cm^{-2}, the transition voltage shifts by approximately −35 V from the pristine state.

We precisely quantified the light-induced modulation of logic transitions by performing differential analysis of the VTC curve to extract the first transition voltage, denoted as V_{sensor}, which corresponds to the critical point where the output logic state switches from '1' to '0'. As shown in Fig. 4(e)–(g), this point appears as a pronounced peak in the first derivative of V_{out} with respect to V_{in}, providing a consistent and quantifiable reference for signal encoding.

For each RGB channel, the device was illuminated five times under increasing light power densities ranging from 1 to 20 mW cm^{-2}. These measurements were used to construct a robust mapping between V_{sensor} and the incident light power density, enabling the extraction of distinct voltage responses under different spectral and power conditions. As summarized in Fig. 4(h)–(j), this mapping reveals a monotonic and color-dependent shift of V_{sensor} values, where higher light power densities consistently result in more negative transition voltages. The extracted V_{sensor} values were further converted into pixel brightness levels for red, green, and blue channels, serving as the quantized input signals to a TNN for image classification. This voltage sensing mode[35] allows direct encoding of light power density into discrete corresponding voltage states, enabling the device to function as a suitable optoelectronic sensing unit for color image recognition.[36]

Furthermore, the parallel IGZO ternary-inverter pixel cell has the potential to be scaled into a dense two-dimensional pixel matrix. In a practical array, non-uniform illumination and noise can be mitigated by adjusting the discrete V_{sensor} levels to provide sufficient tolerance margins. Spatial crosstalk can be reduced through a combination of optical isolation and circuit isolation. Limited dynamic range can be addressed *via* device-level optimization to keep the first transition voltage well defined over the intended illumination range. The device statistics summarized in Table S4 further support the robustness of the V_{sensor} mapping against device mismatch and environmental fluctuations. Finally, to realize a complete realistic vision front end, additional peripheral readout circuitry would be required to enable calibration, error correction, and redundancy for reliable operation.

Importantly, the proposed voltage encoding is not limited to the particular standard image dataset classification demonstration shown later. By properly tuning the number and distribution of V_{sensor} discrete levels to balance the feature contrast, it can serve as a general sensing platform and is potentially extendable to high content imaging data such as biomedical microscopy.[37,38]

Fig. 5 illustrates the proposed ternary neural network (TNN) framework integrated with IGZO-based parallel TFPT devices. In Fig. 5a, the RGB pixel values of an input image are translated into voltage matrices, providing an electrical encoding of visual information that is inherently compatible with hardware-level processing. The detailed method for mapping RGB pixel values to voltage matrices is provided in the Experimental section. Fig. 5b shows the IGZO circuit realization, where parallel TFPTs generate three well-defined conductance states corresponding to ternary weights {−1, 0, 1}. These states are experimentally characterized by applying V_{in} of −10 V, −1 V, and 10 V, respectively, with 100 repeated measurements performed for each condition to obtain the averaged current values. The corresponding current–voltage characteristics in Fig. 5c clearly demonstrate the discrete separation of these quantized states, confirming their suitability for ternary computation. Fig. 5d presents the classification results on CIFAR-10. A TNN without voltage mapping, which uses ideal inputs that retain the original RGB values and operates without the proposed optical-to-voltage conversion, achieves an accuracy of 94.68%, whereas a TNN with voltage mapping achieves 90.14%. Although a slight reduction in accuracy is observed, the voltage mapping provides a robust linkage between algorithmic encoding and device-level implementation, thereby ensuring reliable and practical hardware integration. Table S5 compares the proposed architecture with previously reported transistors for ternary logic in quantized neural networks. Unlike prior works that focused only on digit recognition, our approach integrates photosensing and ternary logic in a parallel IGZO device configuration.

In this work, the measured ternary currents ''OFF'', ''Intermediate'', and ''ON'' are logarithmically transformed and normalized to obtain ternary weights of −1, 0, and 1 for network inference. Network-level simulations assume ideal ternary weights and do not explicitly consider non-idealities, because the measured states are well separated for direct hardware quantization; moreover, the measured device data are far smaller than the neural-network scale, so incorporating non-ideal factors into training and inference would not be statistically representative and could be overinterpreted. With sufficiently large-scale statistics, prior studies suggest that device-aware training can improve robustness and accuracy,[39,40] which will be pursued in future work. Scaling to larger architecture requires managing fan-in, fan-out and connectivity to reduce leakage and noise, avoid loading-induced signal loss, and limit latency and additional energy consumption.

A ternary neural network is adopted to match the three current states provided by the parallel IGZO TFPT circuits as ternary synaptic weights, establishing a direct correspondence between hardware and network. This alignment avoids additional multi-bit weight quantization and complex peripheral mapping circuitry, enabling a compact and consistent ternary

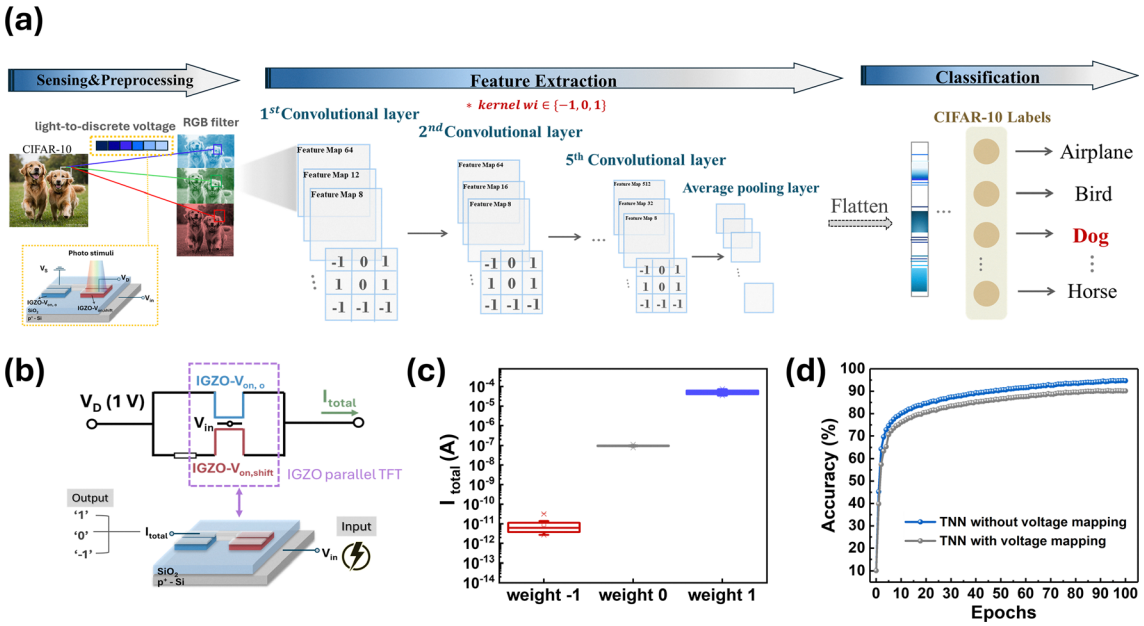

Fig. 5 A schematic of the proposed ternary neural network (TNN) framework integrating the image sensor using the IGZO parallel TFPT device. (a) The input image is separated into three color channels (R, G, B), each of which is encoded as a voltage matrix using an IGZO-based inverter under varying light wavelengths and light power densities. These voltage-encoded 2D matrices serve as the input to the ternary neural network (TNN). (b) Schematic of the IGZO-based ternary circuit using parallel TFPTs, which encodes ternary weight states {−1, 0, 1} through three distinct conductance levels as the hardware basis for ternary convolution operations. (c) *I*–*V* mapping of the equivalent model showing discrete ternary weight states {−1, 0, 1} for direct hardware ternary quantization. (d) Accuracy comparison of TNNs with and without voltage mapping, both exceeding 90%, demonstrating sufficient reliability.

computation framework that integrates sensing–encoding and ternary-weighted inference.

Conclusions

With the convergence of device-level engineering and neuromorphic vision computing, we have demonstrated a compact IGZO-based circuit that enables efficient hardware realization of ternary neural networks (TNNs). By carefully tuning sputtering conditions, the turn-on voltage of IGZO TFPTs was engineered to produce a stable intermediate current state, thereby establishing robust multi-level conduction. This intrinsic electrical characteristic was further exploited to achieve three well-separated current states (OFF, Intermediate, and ON), which were directly mapped to ternary weight values {−1, 0, 1} and served as reliable hardware encoders for TNN kernels. Furthermore, by configuring parallel IGZO TFPTs into a ternary inverter circuit, we realized native three-level inversion logic without relying on an auxiliary CMOS system.

Leveraging the intrinsic photoresponsivity of IGZO, RGB laser illumination on the ternary inverter directly modulates the voltage transfer characteristics, thereby enabling direct RGB pixel value-to-voltage mapping without auxiliary conversion circuitry. Moreover, through the integration of color mapping with ternary *I*–*V* characteristic derived weight quantization in ternary neural networks, the proposed circuit achieves an accuracy exceeding 90% on the CIFAR-10 color image recognition task. These results highlight the potential of IGZO-based ternary circuits as scalable and energy-efficient building blocks for future artificial vision systems, bridging device-level optoelectronics with neuromorphic computation.

Experimental

Fabrication of IGZO parallel phototransistors

Both Al/IGZO–$V_{on,o}$/SiO_2/p^+–Si and Al/IGZO–$V_{on,shift}$/SiO_2/p^+–Si TFPTs were fabricated for the IGZO parallel TFPT configuration. A 100 nm SiO_2 layer as a gate dielectric was thermally grown on a highly P-type doped Si substrate. First, an IGZO (In : Ga : Zn : O = 1 : 1 : 1 : 1.4) active layer was deposited by RF magnetron sputtering with a power of 50 W under a working pressure of 10 mTorr for 4 minutes. Subsequently, the deposited IGZO layer was covered with a shadow mask, and a second sputtering process was carried out using a higher RF power of 70 W under identical conditions to those used in the first process. Finally, Al as the source and drain electrodes (100 nm thick) were evaporated *via* e-beam evaporation through a shadow mask to define the phototransistor channel width (2000 μm) and length (80 μm).

Material characterization

The film thickness and EDS elemental mapping analysis were obtained using a field emission transmission electron microscope ([EM000800] JEOL JEM-2100F Cs STEM), with an acceleration voltage of 200 keV. The absorption spectrum of the IGZO was collected using a UV-vis/NIR spectrophotometer (UH5700, Hitachi). The chemical bonding states of the oxygen atoms in the IGZO films were characterized by X-ray photoelectron spectroscopy (JEOL, JAMP-9500F Auger Electron Spectroscopy).

Electrical and optoelectronic characterization

The electrical and optoelectronic characteristics of the IGZO phototransistor were investigated using Keysight 81150A and B1500A semiconductor parameter analyzers in an ambient atmosphere box. For optical illumination, three laser sources (SDL-405-LM-010T) with wavelengths of 670 nm (R), 520 nm (G), and 405 nm (B) were employed as light sources. The lasers were modulated by the Keysight 81150A to provide fixed wavelengths with variable light power densities, enabling precise on–off switching during measurements. The output power density of the laser sources was monitored using a laser power meter (Model 843-R, Newport).

Color image encoding

To evaluate the sensing capability of our device, we emulated how visual information could be represented by voltage signals. Each CIFAR-10 image was separated into red (R), green (G), and blue (B) colors, with each channel treated as a 2D matrix of 8-bit pixel values ranging from 0 to 255. These values were evenly divided into six intervals, with each interval spanning 42 consecutive pixel values. Each interval was then assigned to one of six light densities (ranging from 0 to 20 mW cm^{-2}) used in our experiments, and the corresponding output voltages were measured for each color channel under illumination conditions.

By replacing the original pixel values in each channel matrix with their corresponding measured voltages, each image was effectively transformed into a voltage matrix, and these values directly reflect the device's photoresponse. In this framework, image information is not sensed as digital data but is instead represented as an optical input, where RGB laser power densities derived from image content are projected onto the device to enable direct optoelectronic encoding.

Structure of ternary neural networks

The proposed model adopts a ResNet-18 backbone, where the convolutional layers are quantized into ternary weights {−1, 0, 1}, according to the formula:

$$q(k) = \begin{cases} +1, & k > \Delta \\ 0, & |k| \le \Delta \\ -1, & k < -\Delta \end{cases}$$

while latent full-precision kernels are retained to facilitate gradient back-propagation. The quantization threshold is given by $\Delta = t \cdot \max(|k|)$, where t is a learnable factor that adapts to the weight distribution.[41] For the input representation, RGB pixel values are first converted into voltage levels through a predefined mapping table. In parallel, ternary weights are expressed as corresponding average current values, providing a device-level abstraction that links algorithmic ternarization with physical signal domains. The ResNet-18 architecture consists of 18 layers (17 hidden layers followed by one fully connected layer) with residual connections, including multiple 3 × 3 convolutional layers, batch normalization, and ReLU activations. The network concludes with an average pooling layer and a fully connected output layer for classifying 32 × 32 pixel RGB images into ten categories in the CIFAR-10 dataset.

Author contributions

Wun-Yun Lin: investigation, conceptualization, data curation, formal analysis, methodology, software, validation, visualization and writing the original draft; Yong-Yi Huang: methodology, validation, editing and writing the revised version; Yu-Chieh Chen: methodology, validation and editing; Chen-Gang Jang: validation; Li-Chung Shih: methodology; Jen-Sue Chen: funding acquisition, supervision, project administration, validation and editing. All authors reviewed the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). The supplementary information includes IGZO parallel TFPT device schematic, optical microscope image, cycle-to-cycle and device-to-device variation in transfer characteristics measured in dark and under illumination, UV-Vis transmittance and the derived optical bandgap of IGZO, ternary inverter characteristics, UPS spectra, XPS spectra, and a comparison table of transistors for ternary logic circuit applied in quantized neural networks. See DOI: https://doi.org/10.1039/d5nh00720h.

Acknowledgements

This work is supported by the National Science and Technology Council, Taiwan under Projects NSTC 113-2124-M-006-008-MY3 and 114-2223-E-006-003 and 114-2923-E-006-008-MY3 and 114-2221-E-006-104-MY3.

Notes and references

1 Y. Wang, Q. Sun, J. Yu, N. Xu, Y. Wei, J. Cho and Z. Wang, *Adv. Funct. Mater.*, 2023, **33**, 26.
2 V. Kornijcuk, H. Lim, I. Kim, J. Park, W. Lee, J. Choi, B. Choi and D. Jeong, *Sci. Rep.*, 2017, **7**, 13.
3 X. Yang, K. Zhu, X. Tang, M. Wang, M. Zhan, N. Lu, J. Kulkarni, D. Pan, Y. Liu and N. Sun, *IEEE J. Solid-State Circuits*, 2023, **58**, 1450–1461.
4 H. Jeong, S. Kim, K. Park, J. Jung and K. Lee, *IEEE Trans. Circuits Syst.*, 2023, **70**, 1739–1743.
5 S. Yin, Z. Jiang, J. Seo and M. Seok, *IEEE J. Solid-State Circuits*, 2020, **55**, 1733–1743.
6 N. Kim, D. Shin, W. Choi, G. Kim and J. Park, *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, **32**, 2925–2938.
7 S. Mittal, *J. Syst. Architect.*, 2020, **104**, 26.
8 C. Kyrkou, *IEEE Trans. Neural Netw. Learn. Syst.*, 2024, 8, DOI: 10.1109/tnnls.2024.3380827.

9 H. Alemdar, V. Leroy, A. Prost-Boucle and F. Pétrot, 2017.
10 Z. Li, X. Huang, L. Xu, Z. Peng, X. Yu, W. Shi, X. He, X. Meng, D. Yang, L. Tong, X. Miao and L. Ye, *Nano Lett.*, 2023, **23**, 11710–11718.
11 A. Laborieux, M. Bocquet, T. Hirtzlin, J.-O. Klein, L. H. Diez, E. Nowak, E. Vianello, J.-M. Portal and D. Querlioz, 2020.
12 X. Wang, C. Dong, Z. Wu and Z. Cheng, *Chin. Phys. B*, 2021, **30**, 12.
13 A. Prost-Boucle, A. Bourge and F. Pétrot, *ACM Trans. Reconfigurable Technol. Syst.*, 2018, **11**, 24.
14 Y. Baek, B. Bae, J. Yang, D. Lee, H. Lee, M. Park, T. Kim, S. Kim, B. Park, G. Yoo and K. Lee, *Adv. Electron. Mater.*, 2023, **9**, 12.
15 Y. Kwon, S. Park, J. Kim, Y. Yoo, S. Lee and J. Cho, *Nat. Commun.*, 2025, **16**, 8.
16 K. Woo, Y. Lee, J. Han, T. Park, Y. Jang and C. Hwang, *Appl. Phys. Lett.*, 2024, **124**, 4.
17 X. Zhu, M. Xi, J. Wang, P. Zhang, Y. Li, X. Luo, L. Bai, X. Chen, L.-M. Peng and Y. Cao, *Sci. Adv.*, 2025, **11**, eadt1909.
18 Z. Deng, Y. Yu, Y. Zhou, J. Zhou, M. Xie, B. Tao, Y. Lai, J. Wen, Z. Fan, X. Liu, D. Zhao, L. Feng, Y. Cheng, C. Huang, W. Yue and W. Huang, *Adv. Mater.*, 2024, **36**, 9.
19 C. Lee, C. Lee, S. Lee, J. Choi, H. Yoo and S. Im, *Nat. Commun.*, 2023, **14**, 11.
20 C. Jo, J. Kim, J. Kwak, S. Kwon, J. Park, J. Kim, G. Park, M. Kim, Y. Kim and S. Park, *Adv. Mater.*, 2022, **34**, 14.
21 B. Jeong, J. Lee, M. Ku, J. Lee, D. Kim, S. Ham, K. Lee, Y. Kim and H. Park, *Nano-Micro Lett.*, 2025, **17**, 24.
22 M. Han and V. Tsukruk, *ACS Nano*, 2023, **17**, 18883–18892.
23 S. Song, C. Choi, J. Ahn, J. Lee, J. Jang, B. Yu, J. Hong, Y. Ryu, Y. Kim and D. Hwang, *InfoMat*, 2024, **6**, 11.
24 S. Lee, Y. Lee, S. Kang, S. Mun, J. Choi and C. Hwang, *ACS Appl. Electron. Mater.*, 2023, **5**, 6686–6696.
25 M. Mativenga, F. Haque, M. Billah and J. Um, *Sci. Rep.*, 2021, **11**, 12.
26 C.-F. Hu, T. Teng and X.-P. Qu, *Solid-State Electron.*, 2019, **152**, 4–10.
27 K. Jang, J. Raja, J. Kim, C. Park, Y.-J. Lee, J. Yang, H. Kim and J. Yi, *Semicond. Sci. Technol.*, 2013, **28**, 085015.
28 K. Hoshino, D. Hong, H. Q. Chiang and J. F. Wager, *IEEE Trans. Electron Devices*, 2009, **56**, 1365–1370.
29 J. H. Park, Y. G. Kim, S. Yoon, S. Hong and H. J. Kim, *ACS Appl. Mater. Interfaces*, 2014, **6**, 21363–21368.
30 X. Wang, Y. Shao, X. Wu, M. Zhang, L. Li, W. Liu, D. Zhang and S. Ding, *RSC Adv.*, 2020, **10**, 3572–3578.
31 A. Sen, H. Park, P. Pujar, A. Bala, H. Cho, N. Liu, S. Gandla and S. Kim, *ACS Nano*, 2022, **16**, 9267–9277.
32 M. Lu, C. Chen and M. Lu, *Phys. Rev. Appl.*, 2016, **6**, 14.
33 J. Yao, N. Xu, S. Deng, J. Chen, J. She, H. Shieh, P. Liu and Y. Huang, *IEEE Trans. Electron Devices*, 2011, **58**, 1121–1126.
34 N. Tiwari, R. Chauhan, P. Liu and H. Shieh, *RSC Adv.*, 2016, **6**, 75693–75698.
35 D. Jayachandran, A. Pannone, M. Das, T. Schranghamer, D. Sen and S. Das, *ACS Nano*, 2022, 13, DOI: 10.1021/acsnano.2c07877.
36 C. Hung, Y. Chiang, Y. Lin, Y. Chiu and W. Chen, *Adv. Sci.*, 2021, **8**, 9.
37 I. Iqbal, I. Ullah, T. Peng, W. Wang and N. Ma, *Eng. Appl. Artif. Intell.*, 2025, **139**, 109573.
38 X. Xu, W. Wang, Y. Liu, J. Backemo, M. Heuchel, W. Wang, Y. Nie, I. Iqbal, K. Kratz, A. Lendlein and N. Ma, *Nat. Mater.*, 2024, **23**, 1748–1758.
39 S. Wu, H. Huang, S. Wang, G. Chen, C. Zhou and D. Yang, *Sci. Adv.*, 2025, **11**, eadr3903.
40 N. Perez-Nieves, V. C. H. Leung, P. L. Dragotti and D. F. M. Goodman, *Nat. Commun.*, 2021, **12**, 5791.
41 F. Li and B. Liu, ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, 1–5, DOI: 10.1109/ICASSP49357.2023.10094626.

