

RESEARCH ARTICLE OPEN ACCESS

Multimode Oxide-Based Optoelectronic Memtransistor for In-Sensor Vision Processing

Min Gu Lee¹ | Geunyoung Kim² | Hwa Young Kim¹ | Do Hoon Kim² | Hanchan Song³ | Huisu Noh¹ | Jungwan Noh⁴ | Seungwoo Lee⁴ | Sang-Hee Ko Park^{1,4}  | Himchan Cho^{1,4} | Kyung Min Kim^{1,4} 

¹Department of Materials Science and Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea | ²Applied Science Research Institute, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea | ³Electronics and Telecommunications Research Institute (ETRI), Daejeon, Republic of Korea | ⁴Graduate School of Semiconductor Technology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea

Correspondence: Himchan Cho (himchan@kaist.ac.kr) | Kyung Min Kim (km.kim@kaist.ac.kr)

Received: 9 April 2026 | **Revised:** 8 June 2026 | **Accepted:** 30 June 2026

Keywords: charge-trap memory | explainable artificial intelligence | in-sensor computing | multimode operation | optoelectronic memtransistor

ABSTRACT

In-sensor processing is a key enabler for edge computing because it allows sensor data to be processed locally, thereby reducing data movement and improving the efficiency of complex workloads. In particular, emerging computationally intensive applications such as explainable artificial intelligence (XAI) further increase the demand for dedicated in-sensor hardware capable of efficient sensor-level processing. Here, we report an optoelectronic memtransistor (OEMT) that integrates a photoresponsive indium zinc oxide (IZO) channel with a NbO_x/AlO_x charge-trapping gate stack, enabling optical sensing, electrical masking, and non-volatile memory within a single three-terminal device architecture. We experimentally demonstrate sensor-level vision explainable artificial intelligence (VXAI) operation using a 3 × 3 OEMT array and validate its scalability through hardware-based simulations. Compared with a conventional sensing–processing pipeline, the OEMT-based framework achieves a 21-fold improvement in system-level energy efficiency, highlighting its strong potential for efficient sensor-level processing.

1 | Introduction

Recently, artificial intelligence (AI) has become an integral part of everyday life, continuously collecting data in real time and delivering useful outputs with minimal latency. In this context, edge computing and, more specifically, in-sensor computing technologies are increasingly needed to process data in real time [1–3]. This is because, in conventional architectures where sensing and processing are physically separated, data movement between the sensor and the computing unit gives rise to data-transfer bottlenecks, high energy consumption, and increased hardware cost.

One representative example that requires in-sensor computing is performing the complex processing needed for explainable artificial intelligence (XAI) at the sensor level. XAI is a computational algorithm that improves the reliability of AI systems by providing explanations for AI outputs [4–6], and is therefore essential in applications where trust in AI decisions is critical, such as autonomous driving, medical diagnostics, and surveillance [7–11]. To implement XAI, the occlusion sensitivity method is commonly used, in which regional importance is estimated by sequentially masking the parts of the input and reevaluating the model output [12–14]. However, this process requires repeated masking and inference, resulting in substantial

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Advanced Functional Materials* published by Wiley-VCH GmbH

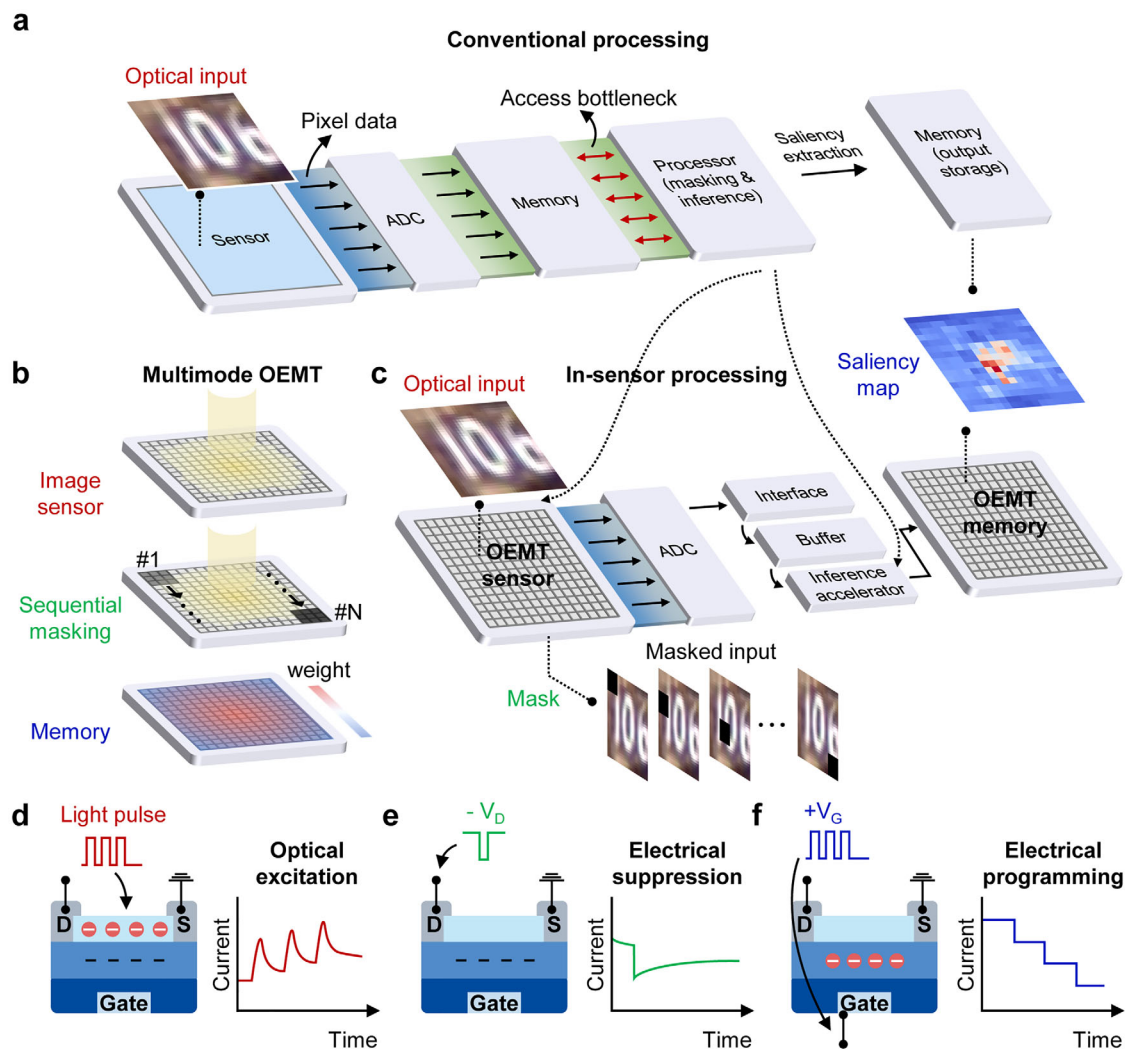


FIGURE 1 | Schematic illustration of the architecture and operational modes of a multimode optoelectronic memtransistor (OEMT)-based vision explainable artificial intelligence (VXAI) system. (a) Schematic of a conventional processing architecture, illustrating the inherent access bottleneck. (b) Conceptual illustration of the multimode OEMT array performing image sensing, sequential masking, and saliency storage within a unified platform. (c) In-sensor processing workflow with sensor-level masking of optical inputs and subsequent saliency map generation. (d–f) Device characteristics corresponding to each operational mode illustrated in (c). Optical excitatory response for image sensing in (d), electrical suppression for masking in (e), and electrical programming for saliency storage in (f).

computational overhead in the conventional hardware framework (Figure 1a).

In 2024, Song et al. demonstrated the feasibility of dedicated hardware for XAI by implementing key XAI processes, including masking, inference, and memory functions, using multiple distinct memristive devices [15]. However, for such a hardware framework to be applied to in-sensor computing, efficient integration with the sensor is essential, and implementing multiple functions within a single device would likely be a more efficient approach, as it can mitigate the data-transfer issues inherent to architectures based on multiple separate devices.

In this study, we propose an oxide-based optoelectronic memtransistor (OEMT) that supports multimode operation for in-sensor computing and demonstrate its utility through a vision explainable artificial intelligence (VXAI) framework as a representative application. Memtransistors are three-terminal devices

that exhibit programmable, history-dependent conductance controlled by gate and drain biases [16–18]. In the OEMT, this bias-dependent operation is further combined with optical sensing capability, enabling the implementation of sensing, masking, and memory functions required for VXAI within a single device (Figure 1b), thereby providing a compact and energy-efficient framework for in-sensor computing (Figure 1c).

The proposed OEMT supports this framework through three distinct operating modes. Under optical stimulation, the device exhibits an excitatory photocurrent response, enabling optical sensing (Figure 1d) [19, 20]. Under drain bias control, the photocurrent can be electrically suppressed, allowing masking operation (Figure 1e). In addition, under gate bias, the device exhibits programmable non-volatile conductance modulation via charge trapping, which enables memory operation for saliency storage (Figure 1f). Leveraging these multimode operations, we experimentally demonstrate sensor-level VXAI operation and

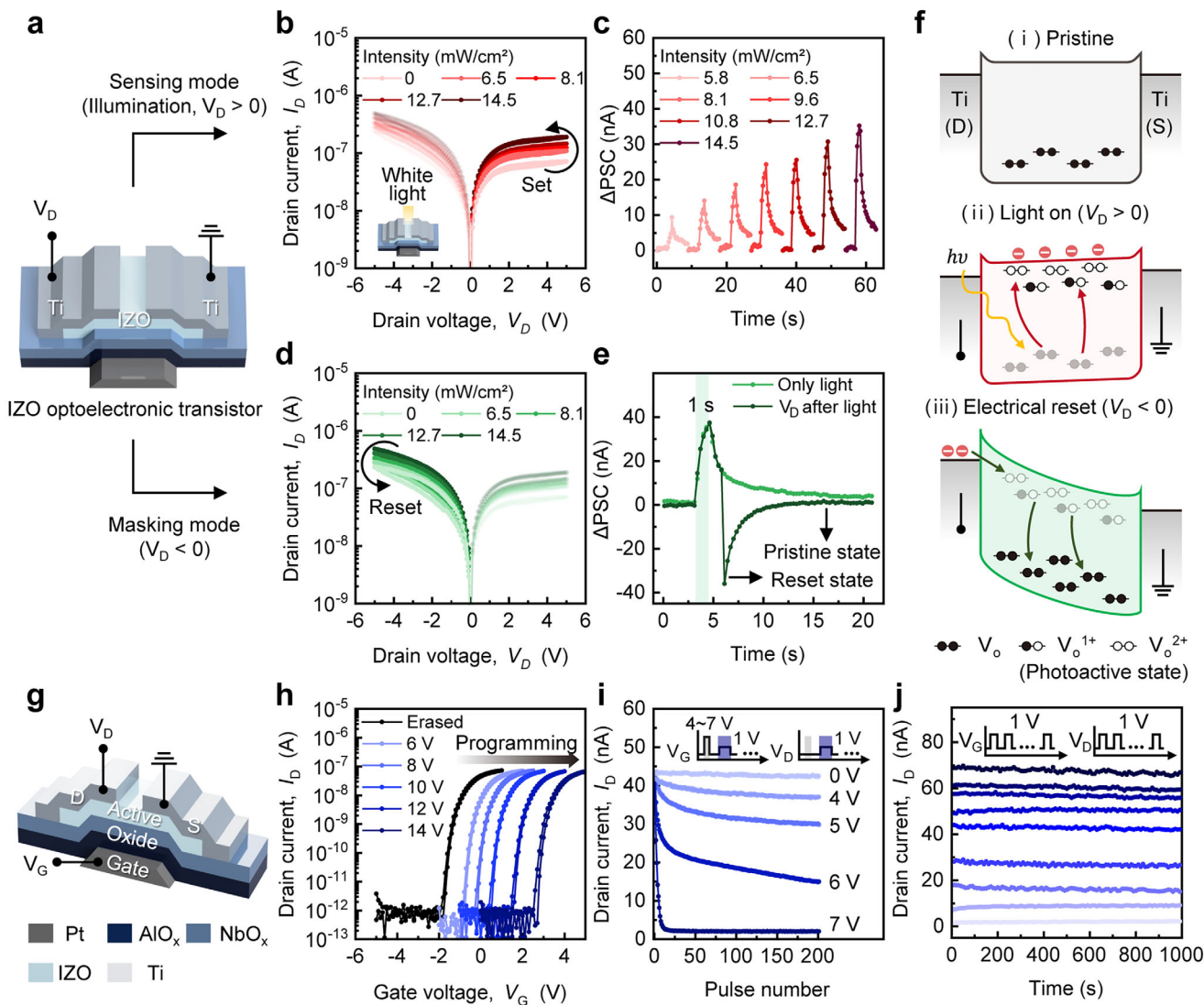


FIGURE 2 | Multimode characteristics of the optoelectronic memtransistor. (a) Schematic illustration of the indium zinc oxide (IZO) optoelectronic transistor showing optical sensing under illumination ($V_D > 0$) and electrical masking ($V_D < 0$). For all illumination, white light was used. (b) I - V curves in the dark and after illumination ($V_D > 0$), showing photo-induced SET behavior. (c) Change in postsynaptic current (ΔPSC) as a function of light intensity. (d) I - V curves in the dark and after illumination ($V_D < 0$), showing electrically induced RESET behavior. (e) Excitatory postsynaptic current (EPSC) and inhibitory postsynaptic current (IPSC)-like responses under optical and electrical stimuli. (f) Band diagrams illustrating the vacancy configurations in the pristine state (i), illuminated state (ii), and electrical reset state (iii). (g) Device structure and measurement configuration for gate-controlled memory operation. (h) Transfer curves under different V_G programming voltages. (i) Depression characteristics with different V_G programming voltages. (j) Retention characteristics of 8 different states.

validate its scalability through hardware-based simulations. These results, together with quantitative benchmarking of energy efficiency and optimized algorithms for large-scale implementation, highlight the potential of the proposed framework for efficient sensor-level XAI processing.

2 | Main

2.1 | Multimode Characteristics of the Optoelectronic Memtransistor

Figure 2a illustrates the OEMT structure and its operating scheme for optical sensing and electrical masking. The detailed device

fabrication process is provided in the Section 4 and Figure S1. Cross-sectional transmission electron microscopy (TEM), energy-dispersive X-ray spectroscopy (EDS) mapping, and atomic force microscopy (AFM) results, which confirm the structural integrity and basic thin-film properties of the OEMT device, are provided in Figure S2. The TEM/EDS analyses confirm the thickness and elemental distribution of the indium zinc oxide (IZO)/ NbO_x / AlO_x /Pt/Ti multilayer stack. In addition, the AFM results of individually deposited IZO, NbO_x , and AlO_x films on a SiO_2 substrate reveal smooth surfaces with average roughness values of 0.20, 0.17, and 0.64 nm, respectively.

The device enables optical sensing, while a negative drain voltage (V_D) suppresses the photoresponse, thereby allowing masking

operation. Figure 2b shows the drain current-drain voltage (I_D – V_D) characteristics measured after 20 s exposure to white light at intensities ranging from 0 to 14.5 mW/cm², confirming the photoresponsivity of the IZO channel. Note that the asymmetric I_D – V_D characteristics can be attributed to drain bias-induced local electrostatic modulation through capacitive coupling under the gate-floating condition. A more detailed discussion is provided in Figure S3.

Regarding the photoresponsivity of the IZO channel, IZO contains a high density of oxygen vacancies, which can be readily ionized under illumination and act as photoactive donors, thereby generating photocarriers and increasing the channel conductance [21, 22]. The ultraviolet–visible absorption spectrum confirms its optical absorption characteristic, while the O 1s X-ray photoelectron spectroscopy (XPS) analysis reveals the presence of oxygen vacancies, supporting the possible presence of photoactive donors (Figure S4). Similar photoresponsive behaviors were also observed in indium gallium zinc oxide and aluminum-doped indium zinc oxide channel devices, although their photoresponsivity was lower than that of the proposed IZO channel device (Figure S5). These results support the selection of IZO as the photoresponsive channel material for the OEMT. Figure 2c shows the change in postsynaptic current (ΔI_D , defined as ΔI_D at $V_D = +1$ V) under different light intensities, where higher light intensity results in larger current responses. During illumination, ΔI_D increases, and upon removal of the light stimulus, it gradually decays, yielding an excitatory postsynaptic current (EPSC)-like transient response [23]. Additionally, paired-pulse facilitation (PPF), a representative form of short-term synaptic plasticity, is observed [24, 25]. More detailed temporal dynamics, including light pulse duration dependence and PPF characteristics, are provided in Figure S6. To further characterize the optoelectronic performance of the OEMT, wavelength dependence, weak-light detection capability, response speed, and energy consumption were evaluated (Figure S7). These key metrics, together with the electrical masking functionality, were compared with previously reported optoelectronic memristive devices in Table S1.

Figure 2d shows the electrical reset behavior, which promotes neutralization of photocarriers under a negative V_D bias, thereby suppressing the increased I_D . Figure 2e compares ΔI_D with and without an electrical reset pulse ($V_D = -12$ V, 10 ms) applied after 1 s of white light illumination at 14.5 mW/cm². The results show that ΔI_D abruptly dropped below its initial value upon applying the V_D pulse and then gradually recovered to its initial state. This electrical inhibitory postsynaptic current (IPSC) behavior serves as the key element for the masking functionality.

The underlying mechanisms of the photonic EPSC and electrical IPSC behaviors are illustrated in Figure 2f. In oxide semiconductors, oxygen vacancy-related defects are well known to act as photoactive donor-like states and contribute to the optical response of the channel (panel i) [26–28]. Upon white light illumination (panel ii), the photoactive donor-like defect states increase the electron carrier concentration and I_D . After the light is turned off, the ionized vacancy states gradually recapture electrons and relax toward the neutral state, leading to a gradual decay of the I_D and a short-term memory effect. In addition, a slowly decaying residual component may persist due to lattice relaxation, leading to long-term memory behavior [29]. When

the reset pulse is applied (panel iii), the electric field promotes the recombination of the charged vacancies into neutral states, resulting in an abrupt drop in I_D . After removing the reset pulse, the I_D gradually recovers as the channel returns toward its equilibrium vacancy distribution.

In addition to the optoelectronic functionality of the IZO channel, the double-layer charge trap layer (CTL) provides non-volatile memory functionality similar to conventional charge trap flash memory. Figure 2g illustrates the measurement setup used to investigate the gate-controlled memory behavior, in which the gate voltage (V_G) and drain voltage (V_D) were biased while the source was grounded. Figure 2h shows transfer curves (I_D – V_G characteristic with $V_D = +1$ V) after applying various V_G programming pulses ranging from +6 to +14 V for 1 ms. After programming, the threshold voltage (V_{th}) shifted more positively with increasing V_G pulse amplitude due to more charge trapping in the CTL [30, 31]. Also, the V_{th} shifted more positively with increasing V_G pulse width (Figure S8). The conduction mechanism analysis indicates that this charge trapping originates from deep trap states located at approximately 1.1 eV in the NbO_x layer (Figure S9).

Without the CTL, the device showed a negligible memory effect, confirming the essential role of the CTL in charge trapping and storage (Figure S10). In addition, when a thinner AlO_x blocking oxide was used, noticeable gate leakage occurred, which hindered the acquisition of stable transfer characteristics (Figure S11).

This non-volatile memory behavior can also be gradually modulated by varying the number of applied V_G pulses, thereby enabling multilevel memory states. Figure 2i shows the progressive depression of I_D as a function of the number of applied programming pulses, which originates from the positive shift in V_{th} . The inset presents the V_G and V_D pulse schemes. During the programming stage, V_G was pulsed from 0 to +7 V for 100 μ s, while V_D was held at 0 V. After each programming step, I_D was measured by applying both V_G and V_D at +1 V for 1 ms. As the number of V_G pulses increases, V_{th} shifts positively, resulting in a gradual decrease in I_D . Figure 2j shows the retention characteristics of multiple I_D levels, confirming stable retention for 1000 s.

Meanwhile, the device does not exhibit electrical erase behavior even under negative V_G , whereas the trapped charges recombine with photogenerated holes under optical illumination (Figure S12). In the charge trap memories, electrical erase is not always efficient when the trapped electrons reside in deep trap states, and the negative V_G condition does not provide sufficient detrapping or hole-injection pathways. In such cases, optical illumination can more effectively erase the trapped charge by generating photogenerated holes that recombine with the trapped electrons. A band-diagram-based explanation of the underlying mechanism, supported by the XPS and ultraviolet photoelectron spectroscopy results, is provided in Figure S13. As a result, the device exhibits a unidirectional memory behavior, in which I_D decreases monotonically from a high to a low state.

The devices exhibited high transfer curve uniformity, with an average V_{th} of –1.87 V and a standard deviation of 0.17 V across 25 devices. In addition, the uniformity of the multimode OEMT

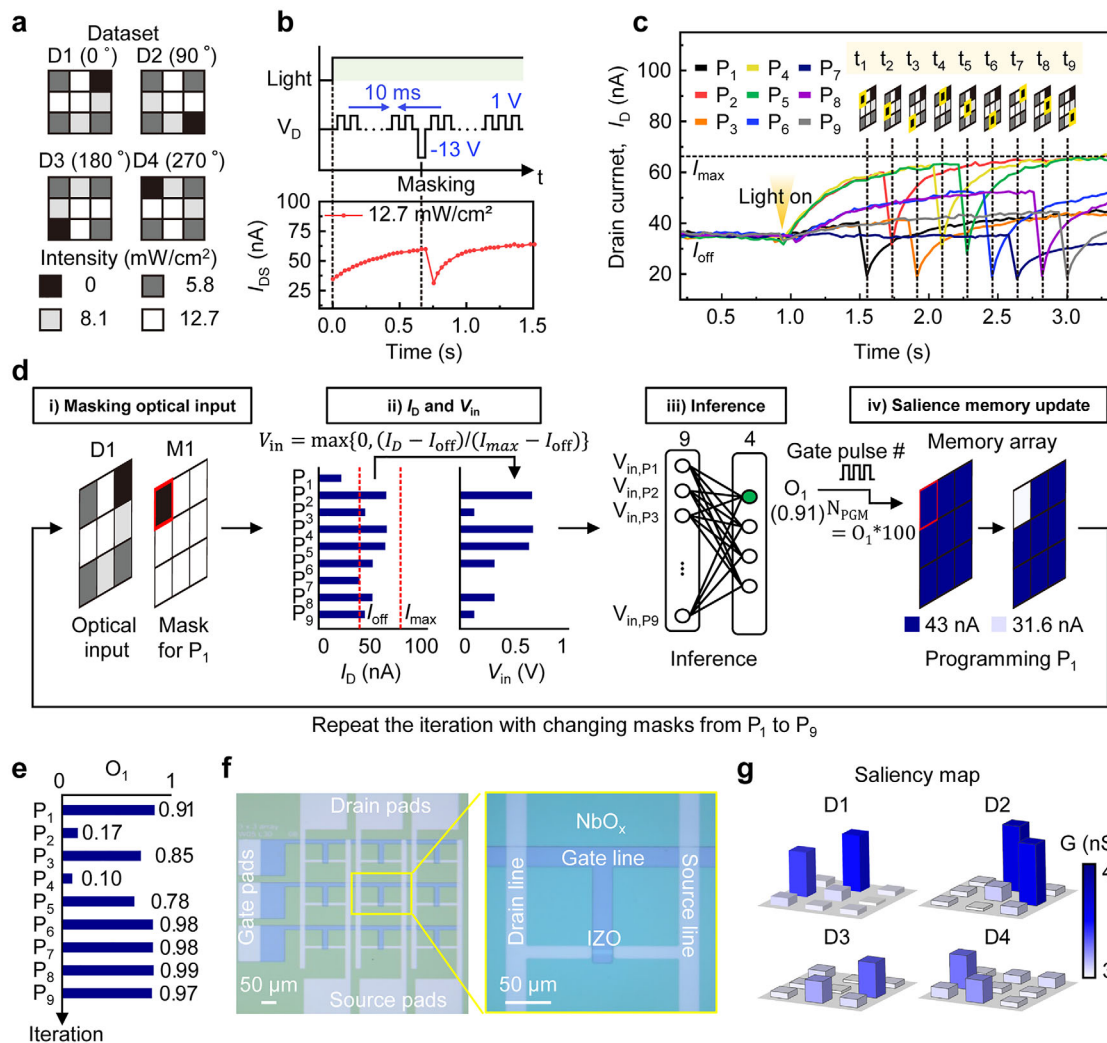


FIGURE 3 | Experimental Vision XAI demonstration. (a) A grayscale testing dataset comprising 4 labels. (b) An example of input and output signals during the optical sensing and electrical masking process: optical and electrical inputs (upper panel), and resulting I_D responses (lower panel). (c) I_D responses from all pixels for the D1 image during the entire process. The inset shows the masking pixels (yellow boxes) at the corresponding masking step (from t_1 to t_9). (d) A workflow of VXAI operation during one iteration with a single masking step, masking P_1 onto the D1 image as an example. (e) O_1 values of the D1 image for different masks from P_1 to P_9 . (f) An optical image of the 3×3 OEMT array (left) and a magnified image of the cell area (right). (g) Obtained saliency maps for D1, D2, D3, and D4 images.

operations was evaluated using 9 individual devices corresponding to a 3×3 array configuration (Figure S14). The extracted EPSC and IPSC responses showed small variations, with coefficient of variation (CV) values of 5.46% and 5.42%, respectively, and the programmed current states under different gate pulse amplitudes remained distinguishable across the 9 devices. The corresponding mean, standard deviation, and CV values are summarized in Table S2. These results support the feasibility of array-level VXAI operation based on uniform optical sensing, electrical masking, and programmable memory functions.

2.2 | Experimental Vision XAI Demonstration

Using the previously described photonic EPSC and electrical IPSC characteristics, we experimentally demonstrated VXAI operation. For this purpose, we used a custom 3×3 grayscale dataset consisting of four images generated by rotating a single pattern

by 0°, 90°, 180°, and 270° (Figure 3a). Each pixel had one of four gray levels (white, light gray, dark gray, and black), corresponding to light intensities of 12.7, 8.1, 5.8, and 0 mW/cm², respectively. A 9×4 single-layer network trained in software achieved 100% classification accuracy on this dataset. Building on this task, we next demonstrate how the OEMT can be utilized to implement VXAI.

Figure 3b shows the light and V_D inputs (upper panel) and I_D outputs (lower panel) during optical sensing and electrical masking processes. At $t = 0$, the light was turned on, and at $t = 0.7$ s, a masking pulse (-13 V, 10 ms) was applied. During this measurement, I_D was periodically read at $V_D = +1$ V using 10 ms pulses at 10 ms intervals. I_D gradually increased under illumination, and abruptly dropped upon application of the masking pulse, and recovered to the illuminated level once the masking pulse was removed. Additional results for four light intensity cases with different masking timings are provided in

Figure S15. In addition, the operational stability of the multimode OEMT was verified using nine individual devices (Figure S16). The devices showed reproducible optical sensing and electrical masking responses over repeated cycles, as well as stable electrical programming and optical erasing operation for memory mode over 100 cycles, supporting the reliability of the multimode switching functions required for VXAI operation.

In occlusion-based XAI, each pixel must be masked at least once to assess its importance. Therefore, for a given 3×3 image, at least nine masking steps are required. Figure 3c shows the I_D responses from all pixels (P_1 to P_9) for the D1 image during the nine masking steps from t_1 to t_9 . The I_D responses for D2, D3, and D4 images are provided in Figure S17. The inset shows the masked images at each masking step with the masked pixel indicated by a yellow box. Upon illumination of the D1 image at 1.0 s, each pixel exhibited an I_D increase in proportion to its assigned light intensity. When the first masking pulse was applied to P_1 at t_1 (at 1.5 s for 10 ms), the P_1 current dropped abruptly. Then, masking pulses were sequentially applied from P_2 to P_9 at 0.18 s intervals, yielding distinct time-dependent I_D profiles for each pixel depending on both light intensity and masking timing.

At each masking step, the saliency map can be updated through the sequence illustrated in Figure 3d. At t_1 , when P_1 of the D1 image was masked (panel i), the I_D values of all pixels were collected (panel ii, left) and linearly converted to the input voltages (V_{in}) for hardware inference (panel ii, right) using $V_{in} = \max\{0, (I_D - I_{off}) / (I_{max} - I_{off})\}$, where I_{off} is the initial current without illumination, and I_{max} is the maximum photocurrent under illumination. These converted inputs were then fed into the neural network for inference (panel iii), and the output corresponding to D1 was then passed through a softmax function. The resulting softmax output (O_1), ranging from 0 to 1, was multiplied by a factor of 100 to convert it to an integer programming pulse number ($N_{PGM} = O_1 \times 100$). In this example, masking P_1 of D1 yielded $O_1 = 0.91$, giving $N_{PGM} = 91$. These programming pulses were then applied to the corresponding saliency memory cell (panel iv). Initially, all memory cells were in the pristine high I_D state (i.e., low V_{th} state). After programming the P_1 cell, its I_D decreased from 43.0 to 31.6 nA. This low I_D state indicates that the pixel is relatively unimportant. This cycle is repeated for all pixels (P_1 to P_9) over t_1 to t_9 , through which the complete saliency map is constructed.

Figure 3e shows O_1 for the D1 image across all nine masking cases. When P_2 and P_4 were masked, the O_1 decreased to 0.17 and 0.10, respectively, indicating that the trained network relies strongly on these pixels to identify the input as D1. This is reasonable because P_2 and P_4 are white pixels that generate strong signals and act as distinctive indicators relative to other patterns. In contrast, the central white pixel contributes little because it is shared across all classes. Also, gray and dark pixels contribute less because they produce weaker signals despite being class-discriminative.

For experimental saliency map generation, we employed a separate 3×3 OEMT array as saliency memory (Figure 3f). The array biasing schemes for programming and reading are provided in Figure S18. Figure 3g shows the final saliency maps for D1, D2, D3, and D4, successfully highlighting the most important pixels (i.e., discriminative white pixels) in each image. Because saliency map

updating is performed solely through electrical programming, the unidirectional programming characteristic of the OEMT does not limit its use for saliency memory operation.

2.3 | Hardware-Aware VXAI Simulation

Next, we tested the feasibility of the VXAI system for a large-scale dataset by simulation. For this simulation, we selected the Street View House Numbers (SVHN) dataset with a resolution of 32×32 pixels [32]. This dataset contains real-world images with multiple digits and background elements, making it well-suited for evaluating VXAI techniques. To adapt the dataset to our VXAI framework, the original images were preprocessed into grayscale with four discrete contrast levels. Figure 4a compares one of the original images (top) and its corresponding preprocessed version (bottom).

For the large-scale data, pixel-level masking is highly inefficient and often fails to provide meaningful interpretations. Instead, patch-level masking (i.e., masking multiple pixels at once) is more effective in terms of both efficiency and interpretability [33, 34]. When patch-level masking is adopted, it is crucial to optimize two hyperparameters, namely masking patch size ($m \times m$) and the masking interval (t_{INTV}), to achieve fast and accurate saliency map generation. A larger m enables faster map generation but may compromise the accuracy of the explanation. Similarly, the longer t_{INTV} allows sufficient recovery of the masked cells, improving the map accuracy, but it is more time-consuming. In this simulation, after thorough optimization (Figure S19), an optimized t_{INTV} of 0.18 s was employed with a 4×4 patch size.

Figure 4b presents a flowchart of one VXAI iteration. Here, the preprocessed input image in Figure 5a (labeled “5” but containing some noise) was used as an optical input. Given the total array size of 32×32 , all pixels could be sequentially masked over 64 iterations with a 4×4 patch size. Figure 4c shows the masked inputs at the sixth, 14th, 22nd, and 30th masking iterations, with the masking region and its scanning direction highlighted by a yellow box and an arrow, respectively (more details on the modeling of photocurrent dynamics for the large-scale simulations are provided in Figure S20 and Table S3). They show that the previously masked pixels retain slightly lower current levels than the original outputs, as they have not yet fully recovered. Nevertheless, this partial recovery has minimal impact on the final saliency map results.

At every iteration, the output currents were converted to the input values and fed to a pre-trained multilayer perceptron (MLP) with a $1024 \times 400 \times 200 \times 10$ architecture. Figure 4d presents saliency maps obtained at the 16th, 32nd, 48th, and 64th (final) iterations. More saliency map examples for different input datasets are shown in Figure S21.

Next, we quantitatively evaluated the validity of the generated saliency map using the insertion test. This method measures changes in the probability score (P-score), defined as the softmax output of the neural network inference for the target label, by progressively adding image patches, starting from the most important patches as defined by the saliency map [35]. Figure 4e shows the saliency map for the D1 image, along with the

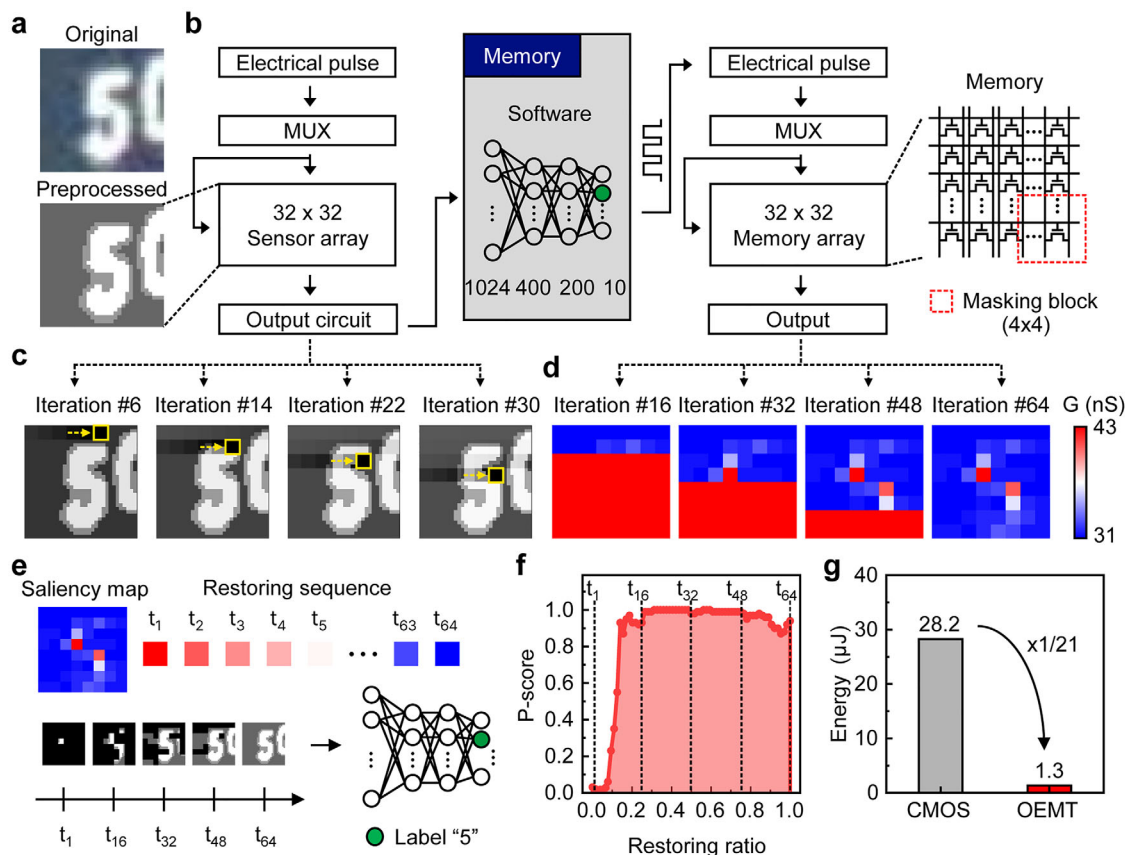


FIGURE 4 | Hardware-aware VXAI simulation. (a) Original image (top) and preprocessed image (bottom). The digit in the center of the image represents the label value. (b) Flowchart illustrating the architecture of the OEMT-based VXAI system. (c) Examples of masked input signals at various iterations. The yellow box represents the current masking region. (d) Generated saliency maps at various iterations. (e) Schematic illustration of the insertion test process. (f) An insertion test result for the test image. (g) Comparison of energy consumption between the conventional complementary metal-oxide-semiconductor (CMOS)-based approach and the OEMT-based one for VXAI operation.

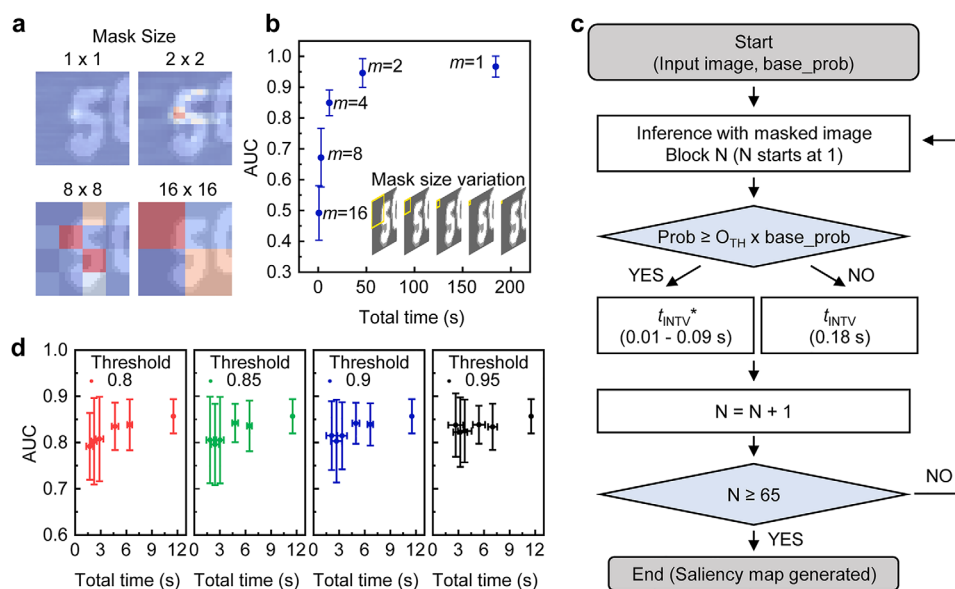


FIGURE 5 | Adaptive scheduling strategy for accelerated VXAI operation. (a) Saliency maps generated with different mask sizes. (b) Comparison of area under the curve (AUC) values and total processing times with different mask sizes. (c) Flowchart of the adaptive scheduling algorithm. (d) Comparison of AUC values with adaptive scheduling under different threshold settings and interval times.

sequentially restored images at the first, 16th, 32nd, 48th, and 64th (final) steps as representative examples. The better the saliency map, the higher the P-score achieved at the earlier insertion steps. Figure 4f shows the insertion test result, where the input image is sequentially restored in the order of the patch's importance. The P-score rapidly increased and stabilized around the 10th insertion, indicating that restoring only 15% of the image was sufficient for the model's prediction. The area under the curve (AUC) of this graph serves as a metric for evaluating the saliency map performance. In this case, the AUC was 0.85, which is significantly higher than the AUC of 0.2 obtained when the insertion test is performed randomly, without using a saliency map (Figure S22). To verify the robustness of the generated saliency maps under realistic sensing conditions, we also evaluated their performance under noisy input conditions. The results confirmed that the proposed VXAI system could reliably capture salient features even under moderate noise levels (Figure S23). Furthermore, the proposed VXAI system exhibits significant energy efficiency at its pre-inference image-processing stage. Assuming a readout time of 10 μ s, which is typical for a row-level readout in a conventional complementary metal-oxide-semiconductor (CMOS) image sensor (CIS), the proposed VXAI system consumes only 4.7% of the energy of a CIS-based digital pipeline for the same task (Figure 4g). Note that although 10 ms read pulses were used in the proof-of-concept VXAI demonstration to ensure stable data acquisition, the readout time can be as short as 20 ns (Figure S24). The detailed energy estimation methodology and readout conditions are provided in Table S4. This substantial reduction originates from its near-sensor integration, which eliminates redundant data transfer and digital processing overhead inherent to traditional CIS pipelines. These results underscore the potential of the VXAI system for real-time and energy-efficient intelligent edge vision applications.

2.4 | Adaptive Scheduling for Accelerating Large-Scale VXAI Operation

Using larger mask patches can reduce the total processing time, but at the cost of reduced explanation accuracy. Therefore, it is essential to optimize the patch size. Figure 5a shows saliency map examples generated with different mask sizes (1×1 , 2×2 , 8×8 , and 16×16). Figure 5b illustrates the trade-off between AUC and processing time, highlighting the inherent difficulty of simultaneously improving both explanation accuracy and VXAI processing speed.

To reduce the processing time without compromising explanation quality, we propose an adaptive scheduling strategy. In this method, if the current masked region is determined to be unimportant based on the softmax output, its restoration becomes less critical, and the subsequent masking iteration can proceed with a reduced interval. Figure 5c shows a flowchart of the adaptive scheduling algorithm. This method requires tuning of two hyperparameters, the softmax output threshold (O_{TH}), which distinguishes unimportant regions from important ones, and the reduced interval time (t_{INTV}^*), which is applied when a region is identified as unimportant. Figure 5d summarizes the adaptive scheduling results under various O_{TH} (0.8–0.95) conditions with a mask size of 4×4 . In each case, t_{INTV}^* was changed from 0.01 to

0.09 s, resulting in the reduced total processing time. For all cases, adaptive scheduling significantly reduced the processing time. Notably, when $O_{TH} = 0.95$, the processing time was substantially reduced from 11.7 to 2.7 s while maintaining a comparable AUC of 0.83 relative to the original 0.85. The results suggest that an appropriate algorithm design can help mitigate the hardware limitations when using the OEMT for the VXAI operations. This highlights the need for further research on hardware-algorithm co-design approaches to fully exploit VXAI technology.

3 | Conclusion

In this work, we propose a sensor-level VXAI hardware framework based on an oxide-based OEMT that integrates optical sensing, electrical masking, and programmable memory within a three-terminal device architecture. Optical stimulation enables sensing with volatile short-term memory, while electrical biasing enables rapid photoresponse suppression for masking as well as gate-programmable non-volatile saliency map storage. These multimodal operations are experimentally validated and further supported by large-scale simulations. Compared with a conventional CIS-based digital pipeline, the proposed framework achieves an approximately 21-fold improvement in system-level energy efficiency. Adaptive scheduling further accelerates VXAI execution by up to 4.3-fold. Moreover, these advantages are realized on a single-device platform, enabling a compact and integrated hardware architecture with potential benefits in fabrication efficiency and system scalability.

Future work will focus on co-optimizing device design and system-level operation to further reduce operating voltages and enhance photoresponse speed and spectral coverage. These advances are expected to extend the framework toward high-speed and multispectral VXAI, providing a scalable pathway to compact, energy-efficient hardware for sensor-level interpretability.

4 | Experimental Section

4.1 | Device Fabrication

The device consists of a Ti/IZO/NbO_x/AlO_x/Pt structure. The device was fabricated on a SiO₂/Si substrate. After gate electrode pattern formation for a lift-off process on a SiO₂/Si substrate, a 4 nm-thick Ti adhesion layer and a 40 nm-thick Pt layer were deposited with an e-beam evaporator. On top of that, a 30 nm-thick AlO_x layer was grown by atomic layer deposition at 250 °C using Trimethylaluminum as a metal precursor and H₂O as a reactant. Subsequently, 30 nm-thick NbO_x was deposited by reactive sputtering, using a Nb metal target in Ar/O₂ (25:5) mixture at 3 mTorr. A 10 nm-thick IZO layer was then deposited on the stack by plasma-enhanced atomic layer deposition at 200 °C. The active layer of the device had a width of 5 μ m and a length of 30 μ m. Subsequently, the active layer was patterned by photolithography followed by wet etching. Finally, the source/drain electrode was patterned by a lift-off process with a 40 nm-thick Ti electrode layer followed by a 20 nm-thick Pt passivation layer deposited with an e-beam evaporator.

4.2 | Electrical Characterization

Current–voltage (I – V) and transfer curves were obtained using a Keithley 4200A-SCS semiconductor parameter analyzer. The pulse measurements were also carried out using a Keithley 4200A-SCS with pulse measuring units (Tektronix 4225-PMU). White light illumination was applied with a lamp power supply (ICN20L-048, MSTECH).

4.3 | Neural Network Simulation

We employed an MLP with an architecture of 1024-400-200-10. The network consists of three fully connected layers, with Batch Normalization applied to the first and second layers to improve training stability. Rectified linear unit activation functions were applied in the hidden layers. The model was trained on the SVHN dataset, which contains 73 257 training images and 26 032 test images. All input images were converted to grayscale, resized to 32×32 pixels, and quantized to four discrete intensity levels before being fed into the network. Training was performed using a cross-entropy loss function and optimized with the Adam optimizer at a learning rate of 0.0001 for 15 epochs and a batch size of 32. To implement occlusion sensitivity, we sequentially occluded some pixels in the test image and used them as input to a pre-trained network.

4.4 | System Energy Consumption Estimation

The energy consumption of the proposed VXAI and the conventional CIS-based digital pipelines was estimated for the pre-inference front-end image-processing stage. In both cases, the total energy was calculated as the sum of the sensor biasing, analog-to-digital conversion, and subsequent data handling, with the specific contributions determined by the underlying system architecture. For the VXAI sensor unit (OEMT-based), the sensing and reset energies were obtained as $E_{\text{sense}} = V_{\text{read}} \times I_{\text{total}} \times t_{\text{read}} \times N_{\text{steps}}$ and $E_{\text{reset}} = V_{\text{reset}} \times I_{\text{reset}} \times t_{\text{reset}} \times N_{\text{cells, reset}}$, respectively, using experimentally validated readout conditions. The energy consumption of the sensor readout and interface circuits was estimated based on a 130 nm Si CMOS technology node. The data-handling energy was modeled as on-chip static random-access memory buffering in the near-sensor VXAI path, whereas it was modeled as microcontroller unit-based processing in the conventional CIS baseline.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) (Grant Nos. RS-2023-00216619, RS-2023-NR077077, RS-2023-00216992, RS-2026-25542514, and RS-2025-02433006).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. F. Zhou and Y. Chai, “Near-Sensor and In-Sensor Computing,” *Nature Electronics* 3 (2020): 664–671, <https://doi.org/10.1038/s41928-020-00501-9>.
2. T. Wan, B. Shao, S. Ma, Y. Zhou, Q. Li, and Y. Chai, “In-Sensor Computing: Materials, Devices, and Integration Technologies,” *Advanced Materials* 35 (2023): 2203830.
3. Y. Baek, B. Bae, H. Shin, et al., “Edge Intelligence Through In-Sensor and Near-Sensor Computing for the Artificial Intelligence of Things,” *Npj Unconventional Computing* 2 (2025): 25.
4. M. T. Ribeiro, S. Singh, and C. Guestrin, “Why Should I Trust You?” Explaining the Predictions of any Classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Digital Library, 2016), 1135–1144.
5. Z. C. Lipton, “The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery,” *Queue* 16 (2016): 31–57.
6. F. Doshi-Velez and B. Kim, “Towards A Rigorous Science of Interpretable Machine Learning,” preprint, bioRxiv, February 28, 2017, <https://doi.org/10.48550/arXiv.1702.08608>.
7. R. Courtland, “The Bias Detectives as Machine Learning Infiltrates Society, Scientists Grapple With How to Make Algorithms Fair,” *Nature* 558 (2018): 357–360, <https://doi.org/10.1038/d41586-018-05469-3>.
8. J. Zou and L. Schiebinger, “AI can be Sexist and Racist — it’s Time to Make it Fair,” *Nature* 559 (2018): 324–326, <https://doi.org/10.1038/d41586-018-05707-8>.
9. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations,” *Science* 366 (2019): 447–453, <https://doi.org/10.1126/science.aax2342>.
10. J. Wiens and E. S. Shenoy, “Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology,” *Clinical Infectious Diseases* 66 (2018): 149–153, <https://doi.org/10.1093/cid/cix731>.
11. J. Buolamwini and T. Gebru, “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification,” in *Conference on Fairness, Accountability and Transparency* 81 (PMLR, 2018): 77.
12. M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks,” in *European Conference on Computer Vision* (Springer, 2014), 818–833.
13. R. C. Fong and A. Vedaldi, “Interpretable Explanations of Black Boxes by Meaningful Perturbation,” in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2017), 3429–3437.
14. S. Rajaraman, K. Silamut, M. A. Hossain, et al., “Understanding the Learned Behavior of Customized Convolutional Neural Networks Toward Malaria Parasite Detection in Thin Blood Smear Images,” *Journal of Medical Imaging* 5 (2018): 034501–034501, <https://doi.org/10.1117/1.JMI.5.3.034501>.
15. H. Song, W. Park, G. Kim, et al., “Memristive Explainable Artificial Intelligence Hardware,” *Advanced Materials* 36 (2024): 2400977, <https://doi.org/10.1002/adma.202400977>.
16. V. K. Sangwan, H.-S. Lee, H. Bergeron, et al., “Multi-Terminal Memtransistors From Polycrystalline Monolayer Molybdenum Disulfide,” *Nature* 554 (2018): 500–504, <https://doi.org/10.1038/nature25747>.
17. X. Yan, J. H. Qian, V. K. Sangwan, and M. C. Hersam, “Progress and Challenges for Memtransistors in Neuromorphic Circuits and Systems,” *Advanced Materials* 34 (2022): 2108025, <https://doi.org/10.1002/adma.202108025>.
18. J. Bae, J. Won, and W. Shim, “The Rise of Memtransistors for Neuromorphic Hardware and In-Memory Computing,” *Nano Energy* 126 (2024): 109646, <https://doi.org/10.1016/j.nanoen.2024.109646>.
19. S.-E. Ahn, S. Park, T. Kim, J. Park, and S. Jeon, “Photoresponse of an Oxide Semiconductor Photosensor,” *Journal of Vacuum Science & Technology B* 33 (2015): 031205, <https://doi.org/10.1116/1.4916626>.

20. J. Choi, J. Park, K.-H. Lim, et al., "Photosensitivity of InZnO Thin-Film Transistors Using a Solution Process," *Applied Physics Letters* 109 (2016): 132105, <https://doi.org/10.1063/1.4963881>.
21. H. Chen, J. Ren, J. Sun, et al., "Photoresponse Design in Metal Oxide Semiconductor TFTs Toward Diverse Applications: Display Drivers, Photodetectors, and Optoelectronic Synapses," *ACS Applied Materials & Interfaces* 17 (2025): 8727–8736, <https://doi.org/10.1021/acsami.5c00152>.
22. P. T. Liu, D. B. Ruan, X. Y. Yeh, Y. C. Chiu, G. T. Zheng, and S. M. Sze, "Highly Responsive Blue Light Sensor With Amorphous Indium-Zinc-Oxide Thin-Film Transistor Based Architecture," *Scientific Reports* 8 (2018): 8153, <https://doi.org/10.1038/s41598-018-26580-5>.
23. Y. Wang, Z. Lv, J. Chen, et al., "Photonic Synapses Based on Inorganic Perovskite Quantum Dots for Neuromorphic Computing," *Advanced Materials* 30 (2018): 1802883, <https://doi.org/10.1002/adma.201802883>.
24. Y. H. Liu, L. Q. Zhu, P. Feng, Y. Shi, and Q. Wan, "Freestanding Artificial Synapses Based on Laterally Proton-Coupled Transistors on Chitosan Membranes," *Advanced Materials* 27 (2015): 5599–5604, <https://doi.org/10.1002/adma.201502719>.
25. C. S. Yang, D. S. Shang, N. Liu, et al., "A Synaptic Transistor Based on Quasi-2D Molybdenum Oxide," *Advanced Materials* 29 (2017): 1700906, <https://doi.org/10.1002/adma.201700906>.
26. M. Lee, W. Lee, S. Choi, et al., "Brain-Inspired Photonic Neuromorphic Devices using Photodynamic Amorphous Oxide Semiconductors and their Persistent Photoconductivity," *Advanced Materials* 29 (2017): 1700951, <https://doi.org/10.1002/adma.201700951>.
27. M.-K. Kim and J.-S. Lee, "Synergistic Improvement of Long-Term Plasticity in Photonic Synapses Using Ferroelectric Polarization in Hafnia-Based Oxide-Semiconductor Transistors," *Advanced Materials* 32 (2020): 1907826, <https://doi.org/10.1002/adma.201907826>.
28. A. Nathan, S. Lee, S. Jeon, and J. Robertson, "Amorphous Oxide Semiconductor TFTs for Displays and Imaging," *Journal of Display Technology* 10 (2014): 917–927, <https://doi.org/10.1109/JDT.2013.2292580>.
29. S. E. Ahn, I. Song, S. Jeon, et al., "Metal Oxide Thin Film Phototransistor for Remote Touch Interactive Displays," *Advanced Materials* 24 (2012): 2631–2636, <https://doi.org/10.1002/adma.201200293>.
30. G. Kim, S. Son, H. Song, et al., "Retention Secured Nonlinear and Self-Rectifying Analog Charge Trap Memristor for Energy-Efficient Neuromorphic Hardware," *Advanced Science* 10 (2023): 2205654, <https://doi.org/10.1002/advs.202205654>.
31. Y. Lee, H. Rhee, G. Kim, et al., "Flexible Self-Rectifying Synapse Array for Energy-Efficient Edge Multiplication in Electrocardiogram Diagnosis," *Nature Communications* 16 (2025): 4312, <https://doi.org/10.1038/s41467-025-59589-2>.
32. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, *Reading Digits in Natural Images with Unsupervised Feature Learning* (Stanford University, 2011).
33. T. Huber, B. Limmer, and E. André, "Benchmarking Perturbation-Based Saliency Maps for Explaining Atari Agents," *Frontiers in Artificial Intelligence* 5 (2022): 903875, <https://doi.org/10.3389/frai.2022.903875>.
34. J. Sun, W. Shi, F. O. Giuste, Y. S. Vaghani, L. Tang, and M. D. Wang, "COVID-19 X-Ray Image Analysis Case Study," *Scientific Reports* 13 (2023): 19488, <https://doi.org/10.1038/s41598-023-46493-2>.
35. V. Petsiuk, A. Das, and K. Saenko, "RISE: Randomized Input Sampling for Explanation of Black-Box Models," in *Proceedings of the British Machine Vision Conference (BMVC, 2018)*, 151.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting File: adfm77205-sup-0001-SupMat.docx.