ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

Hetero-Integrated InGaAs Photodiode and Oxide Memristor-Based Artificial Optical Nerve for In-Sensor NIR Image Processing

*Byungjoon Bae, Minseong Park, Doeon Lee, Inbo Sim, and Kyusang Lee**

In-sensor computing is an emerging architectural paradigm that fuses data acquisition and processing within a sensory domain. The integration of multiple functions into a single domain reduces the system footprint while it minimizes the energy and time for data transfer between sensory and computing units. However, it is challenging for a simple and compact image sensor array to achieve both sensing and computing in each pixel. Here, this work demonstrates a focal plane array with a heterogeneously integrated one-photodiode one-resistor (1P-1R)-based artificial optical neuron that emulates the sensing, computing, and memorization of a biological retina system. This work employs an InGaAs photodiode featuring a high responsivity and a broad spectrum that covers near-infrared (NIR) signals and employs an HfO_2 memristor as the artificial synapse to achieve the computing/memorization in an analog domain. Using the fabricated focal plane array integrated with an artificial neural network, this work performs in-sensor image identification of finger veins driven by NIR light illumination (≈84 % accuracy). The proposed in-sensor image computing architecture that broadly covers the NIR spectrum offers widespread application of focal plane array for computer vision, neuromorphic computing, biomedical engineering, etc.

1. Introduction

Image sensors generate information from the scene in the form of electrical signals.[1–3] The converted signals are transmitted to the post-memory and computing units for targeted image processing. However, the skyrocketing number of sensors and their numerous applications generate a huge amount of information that burdens the data transfer and processing between the sensory and computing nodes in terms of energy consumption, time delay, and system footprint.[4] To alleviate the issue, in-sensor computing has been proposed that can fuse the data acquisition and computing units within a sensory domain.[5] In-sensor computing enables simultaneous sensing and computing on a single chip, reducing data transportation between the building blocks and minimizing the overall system footprint. However, because sensing, memory, and computing functionalities need to be performed in each pixel, in-sensor computing usually requires complicated backplane circuitry compared to the conventional active-matrix image sensor array that detects and stores optical signals in the two-dimensional (2D) structured pixels.[5]

In particular, conventional computing units consist of multiple electronic components to effectively process the acquired signals. The complementary-metal-oxide semiconductor (CMOS) technology-based architectures require more than five transistors even for simple arithmetic computation,[6] which significantly increases the complexity of backplane circuitry for in-sensor computing.[5] Unlike the CMOS architectures that process the signals in the digital domain, emerging neuromorphic computing allows in-memory processing in the analog domain, reducing the energy consumption, data processing time, and footprint of the computing devices.[7–9] One key component of neuromorphic computing is a non-volatile memristor (NVM) that can store information as a resistance of the active medium.[10,11] The stored information as an analog form (resistance) is deployable to realize the multiply-accumulate (MAC) operation via Ohm's and Kirchhoff's laws. The general mechanisms of resistive switching are the formation (SET) and rupture (RESET) of metal cations or oxygen vacancies. The stochastic distribution of atomic transitions allows the analog resistive switching of the NVM.[12] Based on NVM as a pixel, the memristor crossbar array can process matrix-form information (images), by using an electrical signal as an input. Such a memristor array architecture can be realized for in-sensor neuromorphic vision computing by employing photo-responsive NVMs.[1,13] However, most of the photo-responsive NVMs feature a three-terminal structure, requiring multiple interconnections and a transport channel for each pixel.

B. Bae, M. Park, D. Lee, I. Sim, K. Lee
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA 22904, USA
E-mail: kyusang@virginia.edu
K. Lee
Department of Material Science and Engineering
University of Virginia
Charlottesville, VA 22904, USA
The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adom.202201905.

DOI: 10.1002/adom.202201905

Here, we demonstrated a heterogeneously integrated one-photodiode one-resistor (1P-1R) array for in-sensor vision computing that covers the near-infrared (NIR) spectrum. The NIR imaging provides a wide range of intriguing applications that are generally unachievable by human eyes, including time-of-flight ranging,[14] night vision,[15] agronomical investigations,[16] human-computer interactions,[17] vein identification[18], *etc*. Among various III-V compound semiconductors, InGaAs is one of the most promising candidates as a NIR sensor[19] due to its high sensitivity to the NIR spectrum.[20–22] Therefore, we have employed the InGaAs p-i-n diode as a NIR sensor integrated with an artificial synapse for in-sensor computing.

We have also accommodated a HfO_2 oxygen vacancy-based NVM as an artificial synapse due to the high resistive switching endurance and reliability of the HfO_2.[23] The fabricated 1P-1R array architecture features in-sensor computing and memorization functionalities, which emulate the vision processing of the biological neural retina. We have also demonstrated the potential applications of the 1P-1R artificial retina, including Modified National Institute of Standards and Technology (MNIST) image classification and vein identification.[24,25]

2. Results

Here, the functionality of the designed artificial optical nerve is mainly inspired by the biochemical process of photoreceptors that modulate the response to both strong and dim light in human retinal cells.[26–30] To achieve light modulation using artificial optical nerve, we have fabricated an InGaAs photodiode focal plane array integrated with an HfO_2 memristor (1P-1R) that can operate in either in-sensor computing or low-level instantaneous memorization modes. **Figure 1**a,b shows block diagrams of the 1P-1R structure describing the procedure of the two modes. In mode 1 (Figure 1a), the 1P-1R serves as a low-level in-sensor computing system. The output current from each photodiode is tuned in conjunction with the resistance of NVM, emulating the fore-segment of human retinal components. Instead of tuning the photo-response of the photodiode, the synaptic weights are emulated by the NVMs. In other words, the applied bias voltage is divided between the serial photodiode and the NVM. When the NVM is programmed (low-resistance state), more bias is applied to the photodiode to generate more photocurrent in the pixel than under the high-resistance state of the NVM. By doing so, the incident optical image on the 1P-1R array can be preprocessed through the programmed NVM units for image classification, which mimics the signal encoding capabilities of a human vision system.

In mode 2 (Figure 1b), the NVM integrated with a photodiode is programmable using the photogenerated current from the photodiode by incident light. Using this property, we demonstrated in-sensor memorization by programming the resistance of NVM in 1P-1R pixels using incident light. Consequently, the hardware and software co-designed NIR in-sensor memorization system connected with an artificial neural network (ANN) has been employed to directly classify the incident image. We have demonstrated vein pattern identification using the finger vein image under the illumination of NIR (≈750–770 nm) light as a potential application for security purposes. Figure 1c shows a schematic illustration of the chip design and optical image (left). The InGaAs photodiode focal plane array is hetero-integrated with an HfO_2 memristor (1P-1R) on an InP substrate (see Experimental Section). The photodiode in both modes serves as a photoreceptor to generate photocurrent and is also used as a selector during the readout process. The NVM is used as a modulator in mode 1, while it is used as a low-level memory in mode 2.

We have first conducted optoelectrical characterization of a single 1P and 1R to confirm its functionality. Before programming the NVM, a forming procedure has been performed in order to create conductive filament in the HfO_2 film (blue line in **Figure 2**a). We have applied the voltage sweeping from 0 to 6 V under the 100 μA current compliance and confirmed the formation of a conductive filament by observing a 5-decade conductance change in the ≈5–6 V region. The NVM has been reset to a high-resistance state (HRS) by applying a sweeping voltage from 0 to −3 V. The red line in Figure 2a indicates the switching ability of NVM between the HRS and low-resistance state (LRS) in the −2 to 2.5 V window. Repeated SET and RESET operations have been performed by applying the positive and negative sweeps of 1 V for setting and -2 V for resetting, respectively, with current compliance varying from 1 to 0.01 mA (Figure 2b). As shown in Figure 2c, the conductance of programmed NVM increases almost identically as the compliance current increases. This implies that the conductance of NVM is able to be set to the desired level using the current compliances.

Moreover, we have verified the endurance of the 1P-1R as shown in Figure 2d. We applied set (>0.5 V/500 μs) and reset (<−0.8 V/500 μs) pulses 150 times in turn to toggle the state of the NVM with the small read (0.2 V/500 μs) pulses. The NVM's endurance is confirmed by the small standard deviation of the Gaussian-fit histograms of reset ($\sigma = 3.583 \times 10^{-6}$) (Figure 2e) and set ($\sigma = 9.273 \times 10^{-6}$) currents (Figure 2f).

The tunability of NVM conductance is one key factor enabling neuromorphic computing. Figure 2 g exhibits the modulation of conductivity increasing and decreasing via 100 repeating potentiation pulses and depression pulses, respectively. The gradually increasing 500 μs width voltage pulse train from 0.6 to 1.05 V was used for the potentiation process. Similarly, a gradually decreasing 500 μs width voltage pulse train from −0.95 to −1.17 V was used for the depression process. With applied pulse trains, the conductivity linearly increased through the potentiation process from 282 μS (Ω^{-1}) to 1.59 mS (Ω^{-1}), while the conductivity linearly decreased through the depression process from 1.52 to 0.30 mS (Ω^{-1}). Since the fabricated HfO_2-based NVM exhibited linear conductance tuning for potentiating and depressing, the device is suitable for facilitating gradient algorithms for linear regression. The conductance of the NVM after setting was measured by giving a reading pulse voltage (0.2 V) every 10 s to confirm the retentivity of the NVM (Figure 2g). The conductance maintained its original value and exhibited only a 0.1 % change during the measurement. Figure 2i shows the current-voltage (I–V) characteristics of the fabricated InGaAs photodiode in the single 1P-1R device under light illumination at a wavelength of $\lambda = 760$ nm with varying light intensities from 0 to 1.738 W m^{-2}. Under the

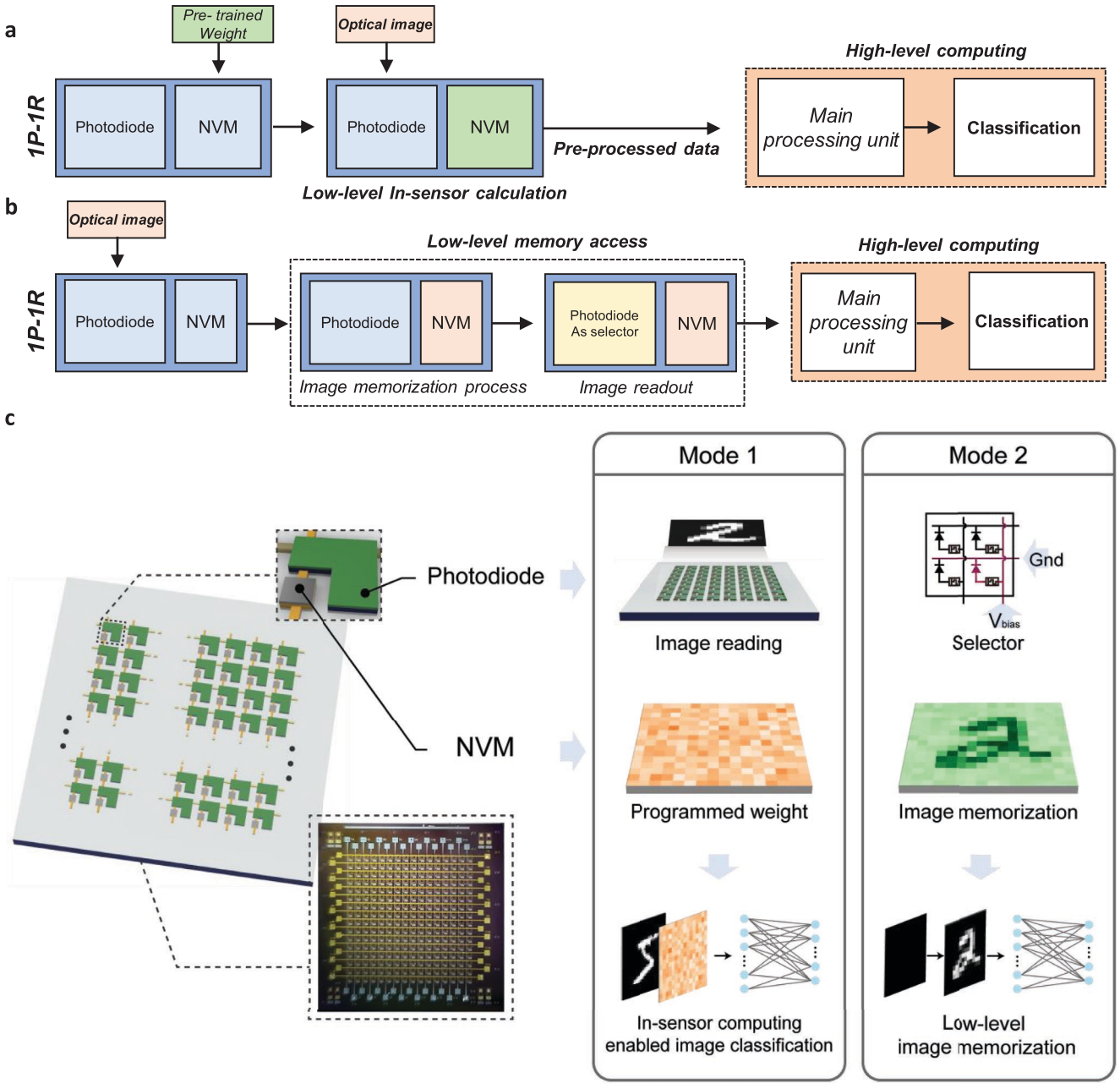

Figure 1. Schematic illustration of overall system. Block diagrams describing in-sensor operation flows of a) mode 1 and b) mode 2. c) Schematic illustrations of fabricated chip and two functionalities of it: i) Integrated non-volatile memristor (NVM) function as modulators, enabling in-sensor computing for improved image classification. Image input is read by a photodiode, and the image readout values are modulated by NVM, which are pre-programmed by the pre-trained weights. The current readout from the one-photodiode one-resistor (1P-1R) device is sent to the artificial neural network (ANN) for classification. ii) Since the NVM integrated into the device is able to be programmed with the input light, the finger vein illuminated with a 770 nm LED light is programmed to the 1P-1R array. With the post-hardware ANN simulated in this work, the system determines whether the current user of the device is identical to the programmed user or not.

constant weight value of the NVM, the photocurrent increased linearly with respect to the incident powers. The external quantum efficiency (EQE) was 55% under the illumination of light at a wavelength of 770 nm without the anti-reflection coating. The extracted responsivity and detectivity are 34.59 A W^{-1}, and 2.54×10^{12} cm $Hz^{1/2}$ W^{-1}, respectively. The measured response speed of the photodiode is 60.0 ns.

Figure 3a shows a schematic illustration of the structure of a single 1P-1R pixel. The single InGaAs photodiode is directly connected with the HfO_2 NVM pixel. The single InGaAs photodiode feature a p-i-n structure composed of 150 nm p-type InGaAs, 2 µm intrinsic InGaAs, and 200 nm of n-type InP. The bottom of the photodiode is connected to the bottom contact (Ti/Pt electrode) of the NVM. To independently access each

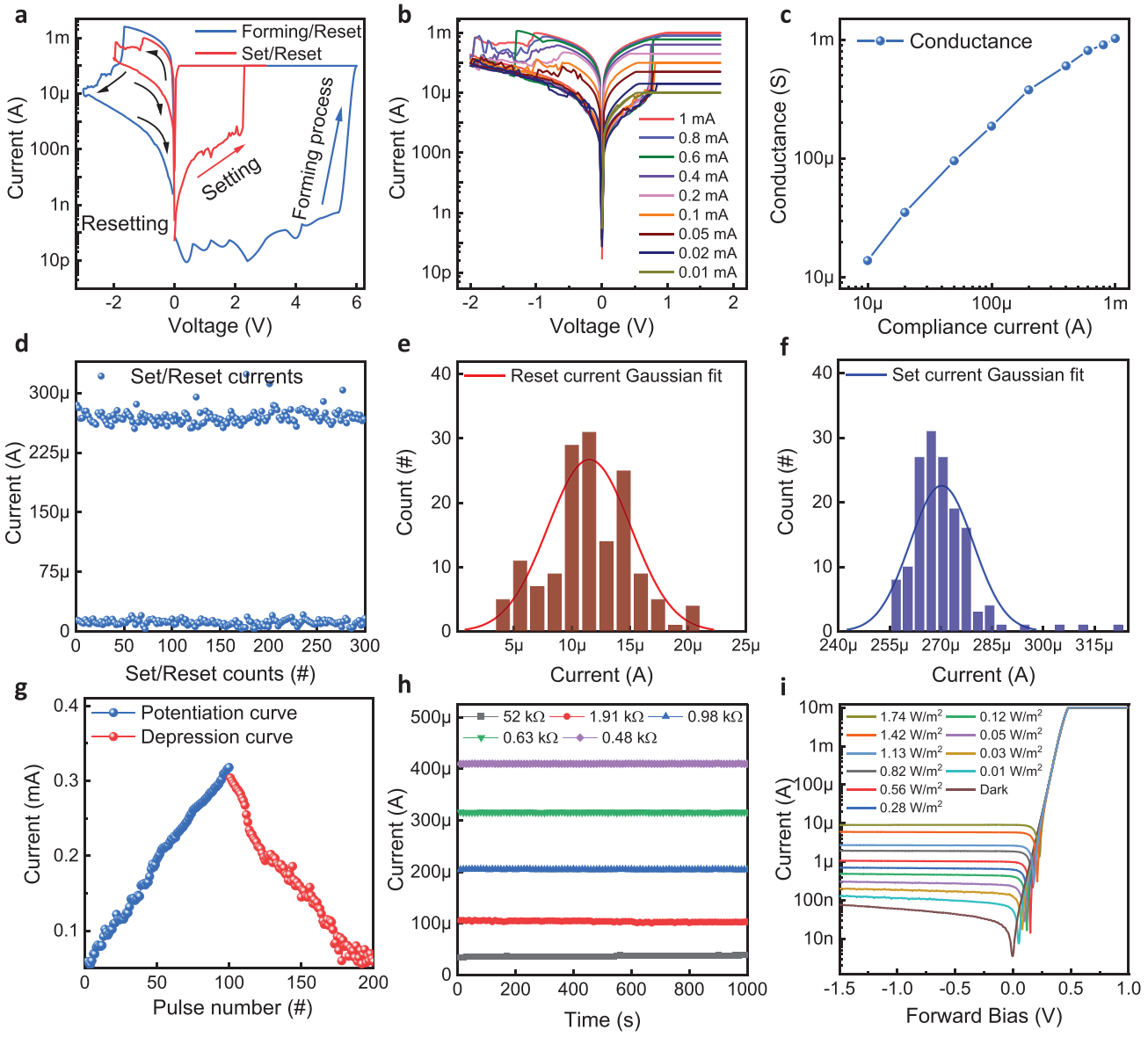

Figure 2. Characterization of integrated non-volatile memristor (NVM). a) Current-voltage (*I–V*) characteristic of single NVM. A conductive filament was formed by sweeping the voltage from 0 to 6 V was used to form the conductive filament. After forming, we used 0 to −2 V sweeping to reset the device (blue line). −2 V to 2.5 V voltage sweep has been done to see the set and reset ability of the NVM (red line). b) Series of set and reset processes with different current compliance. The voltage swept from −2 V to 1.8 V for each process. c) Maximum conductance of NVM in conjunction with each compliance current. The conductance increases as the compliance current increases. d) Readout current after consecutive set and reset process. 1.2 V set voltage pulse was given for the set process, while −1.6 V reset pulse was given for the reset process under the 1 mA current compliance. e,f) Stochastic distribution of reset (Figure 3e) and set (Figure 3f) currents. The distribution shows a Gaussian distribution with low standard deviations ($\sigma < 1 \times 10^{-5}$). g) Potentiation curve and depression curve of single NVM. The potentiation curve and depression curve show decent linearity, which is advantageous for the application of neural network algorithms. h) Retention test of the NVMs set with various conductance. The conductance of NVM was measured every 10 s until the total process time reached 1000 s. i) *I–V* curves of the single photodiode under the dark and various light intensities with a wavelength of 770 nm.

component in the pixel, the individual 1P-1R cell has been designed to have an additional contact between the photodiode and NVM. A magnified view of the single 1P-1R pixel through an optical microscope is depicted in Figure 3b. The inset figure shows HfO_2-based NVM, which has 5 μm × 5 μm dimension. Figure 3c shows a scanning electron microscope (SEM) image of the fabricated 1P-1R 16 × 16 neuromorphic image sensor array. The top electrodes of the NVMs (Ta/Pt electrodes) are shared by the row lines, while the top electrodes of the InGaAs photodiodes (Ti/Pt/Au electrodes) are shared by the column lines. A single 1P-1R pixel has been scaled to a 16 × 16 focal plane array to process the 256-pixel image via two modes. The 1P-1R structure reduces load of data transfer at the sensor-processor interface, simulating biological processes in the

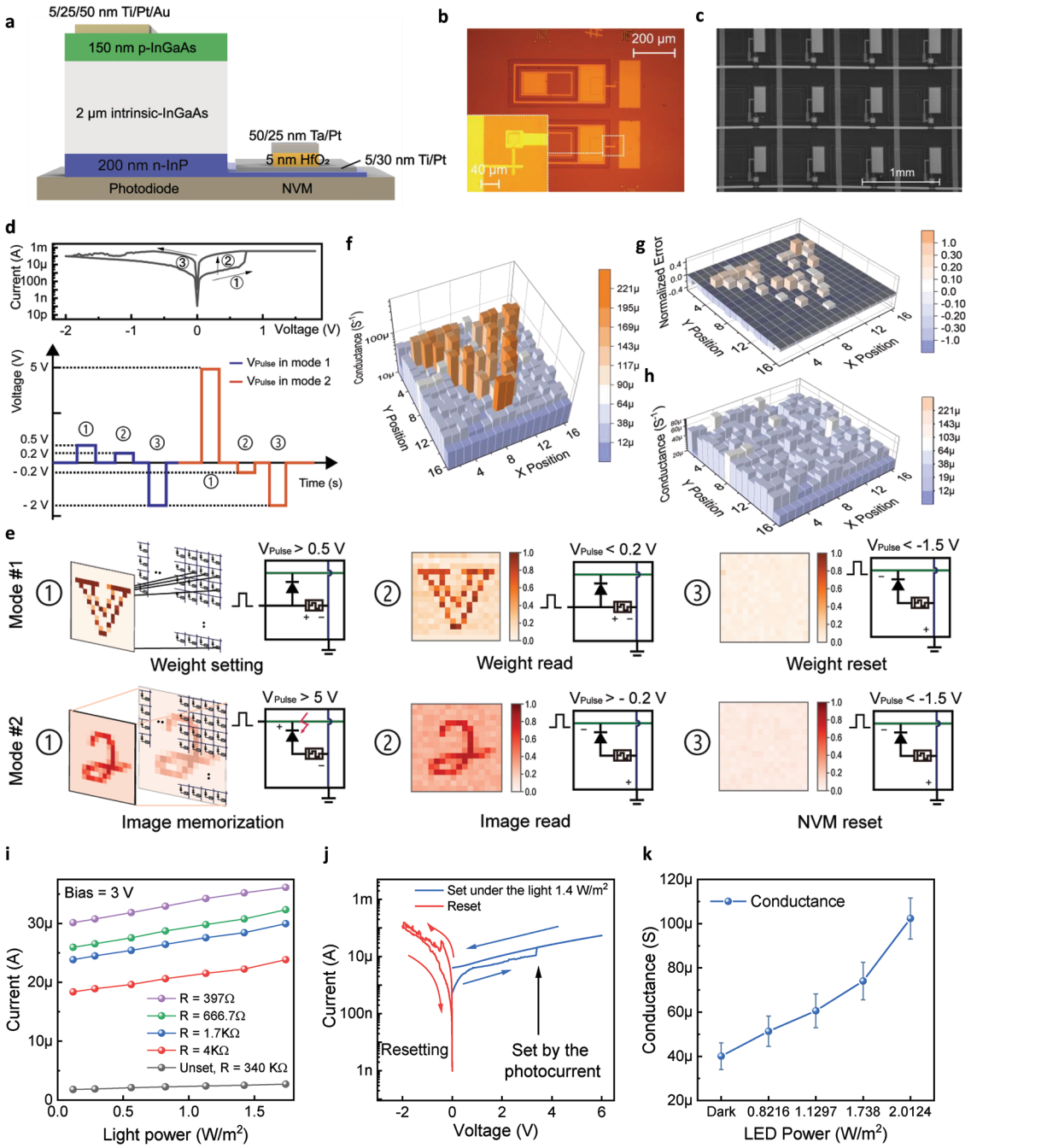

Figure 3. Single 1P-1R and array characterization. a) Schematic illustration of the 1P-1R single-cell device. b) Optical microscope image of the fabricated 1P-1R single cell device. The inset shows a magnified view of the fabricated HfO_2 non-volatile memristor (NVM) region. c) Scanning electron microscopy (SEM) image of fabricated 16 × 16 1P-1R arrays. d) Current-Voltage (*I–V*) curve loop of NVM. Each set/reset/reading process and the corresponding applied voltage pulse are labeled in the figure. e) For mode 1 operation: 1) NVM conductance is set to the target value using potentiation. The programming is done by applying the pulse train (>0.5 V) to the NVM. 2) After setting the NVMs, a readout voltage (0.2 V) was applied to each NVM to confirm the conductance. The result of set conductance is depicted in the left figure. 3) The programmed NVMs were then reset to their original conductivity by applying a negative voltage pulse (greater than –1.5 V). For mode 2 operation: 1) NVM conductance was set under the incident light by applying the voltage pulse (>5 V). 2) After setting the NVMs, a readout voltage (–0.2 V) was applied to the target pixel. 3) The reset process is identical to that of mode 1. f) NVM conductance map. The conductance was set to target value using potentiation, and the target value was converted from normalized value: –1-1 to 10–250 μS. g) Error plot of targeted value and measured value. The error was calculated after normalizing the conductance to ≈0–1. h) Conductance map measured after resetting NVM to original state. i) Single 1P-1R response with various light powers. Under the 3 V bias to 1P-1R, current outputs from the 1P-1R device are affected by the programmed NVM. The higher resistance of NVM causes lower current output to the device. j) Set/reset loop of NVM *I–V* curve with applied sweeping voltage. The NVM is set when applying 3 V to the top of the photodiode. k) Transitions of NVM conductance after setting under various light illumination powers. Under the 5 V bias condition to 1P-1R, the conductance of each NVMs was set higher with the higher intensity of light illumination.

human eye. Figure 3d shows the *I–V* loop curve of the NVM. To operate the NVM in mode 1 and mode 2, pulses with different amplitudes were applied to the NVM.

Figure 3e schematically shows the manipulating process of the NVM in each mode. We first converted the target image to a 256-pixel image and normalized the values from −1 to 1. For mode 1 operation, 1) Each NVM cell was programmed to target conductance values corresponding to the converted image. In each cell, the states were set via potentiation processes by applying voltage pulses of which the amplitude linearly increases from 0.6 (1st pulse) to 1.05 V (98th pulse). 2) Consequently, a readout voltage (0.2 V) was applied to each NVM to measure the conductance. 3) The programmed NVMs were then reset to their original conductivities by applying a negative voltage pulse (>−1.5 V).

For mode 2 operation, 1) NVM conductance was set under the incident light by applying the voltage pulse (>5 V). 2) After setting the NVMs, a readout voltage (−0.2 V) was applied to the target pixel to selectively allow the corresponding photodiode to operate in the forward bias. 3) The reset process is identical to that of mode 1. Figure 3f depicts the conductance measurement result of 1P-1R array after setting to each target value of which the highest and lowest conductivities are 250 μS and 10 μS, which corresponds to normalized pixel values from the image (1 and −1). The error between the targeted value and the measured value is mapped as shown in Figure 3g. The mean absolute value of all errors was small (0.3) for the in-sensor performance. Figure 3h depicts the result of a conductance measurement after resetting the 1P-1R array to HRS.

The output of the photodiode in the reverse bias region is modulated by the integrated NVM since the applied bias divided by the NVM increases as the conductance of the NVM decreases. When a reverse bias of 0 to 5 V is applied, the photodiode operates in a low resistance mode (<100 Ω) compared to NVM (>100 Ω). We programmed the NVM to 5 different resistance levels (397.0 Ω, 666.7 Ω, 1.7 kΩ, 4.0 kΩ, and 340.0 kΩ) and observed the response of the photodiode under the different light illumination intensities. As shown in Figure 3i, the higher current output was observed with the lower resistance level of NVM, which confirms the functionality of 1P-1R system as an artificial optical neuron.

In an integrated 1P-1R system, the photodiode limits the current flowing through the device without the need for current compliance provided by the power/voltage generator. Under the current compliance generated by photocurrent, we confirmed the *I–V* set/reset loop curve of NVM (Figure 3j). Based on the results depicted in Figures 2b,c and 3j, we have programmed the NVM with different current compliance caused by the intensity of light (Figure 3k). Under the 3 V reverse bias condition, the set conductance of NVM has been linearly increased as the intensity of incident light increases since the higher intensity of light induced the higher current. As a result, the HRS state of NVM showed the conductance of 102 μS (±9.27 μS) under the light intensity of 2.01 W m^{-2}, while the LRS state of NVM showed conductance of 51.3 μS (±6.83 μS) under the light intensity of 0.82 W m^{-2}. Utilizing this programmability of NVM depending on the intensity of incident light, the gray scale image was able to be memorized to a 1P-1R array with 5 quantized conductance settings to confirm the feasibility of the system for NIR image identification applications.

We conducted simulations with the fabricated 1P-1R artificial retina, which features two in-sensor modes: the computing mode and the memorization mode. **Figure 4**a shows the schematic illustration of the in-sensor computing mode. For mode 1, we have employed the MNIST digit dataset for the image classification task.[24] The total number of labels is 10, from "0" to "9" digits. We first simulated the ANN which includes the in-sensor computing layer implemented via 1P-1R pixel configuration and the following fully connected layer (FCL). The pre-trained weight matrix of the in-sensor computing layer is mapped as the resistance of the NVMs. To activate the photodiodes, 3 V is applied to the column of the 1P-1R artificial retina. Note that 3 V is a low enough voltage level that the NVMs are not programmed. We have also demonstrated the in-sensor memorization mode for the Finger Vein University Sains Malaysia (FV-USM) dataset that captures vein images of a finger under near-infrared (NIR) illumination.[25] Figure 4b summarizes the memorization process for the vein identification task. The region of interest (ROI) is extracted from the raw image and resized to 30×10 pixels. The vein pattern is optically mapped to the memristor array, driven by $V_{bias,high}$ to activate the photodiode array. The non-volatile feature of the memristor maintains the illuminated vein information, and the readout resistance matrix is applied to the post-ANN.

Unlike a conventional approach that uses the simultaneous biasing of 3 V, we improved the classification performance by employing the sequential biasing of 3 V to each row, which increases the dimensionality of the output (Figure 4c) by implementing element-wise multiplication. The sequential biasing scheme exhibits up to 82% classification accuracy, while the simultaneous biasing approach shows an accuracy of only up to 43% during 200 epochs. The pre-trained 1P-1R artificial retina is also robust enough to perform the classification tasks using noised images, as shown in Figure 4d,e. Both the software-based and device-based simulation results correctly identified the noised images (for more details, see Experimental Section). The energy consumption of in-sensor operation (1 cycle of write-read-erase-read) is calculated to be 187.5 μW, when the pulse 500 μs with a 50% duty cycle is applied. With respect to computing time, the required time to read the result of in-sensor computing is: T_{read} = (500 μs)/0.5 × 16 = 16 ms. According to the calculation, the consumed energy and the processing time are dependent on the employed pulse width; therefore, they can further be minimized by employing the short pulse for the set/read/erase operations.

In mode 2, for 20 labels (20 people), the classification accuracy of the device-based simulation exhibits up to 84% during 200 epochs (Figure 4f). The weight values of the hidden layer have been trained, originating from the initial Gaussian distribution (Figure 4g). We have also performed the vein identification task for various numbers of labels, as shown in Figure 4h. The smaller 10-label dataset exhibits the highest classification accuracy (averaged from the 151st to 200th epoch). To compare this classification capability, we define the classifiability $= \frac{\text{Number of labels which test data is mostly classified}}{\text{Total number of labels}}$. The confusion matrix for 20 labels (Figure 4i) exhibits 4 correct answers ("1") and 2 incorrect answers ("2") for "1" true labels. In this case, the "1" label is classified as a "classifiable" label.

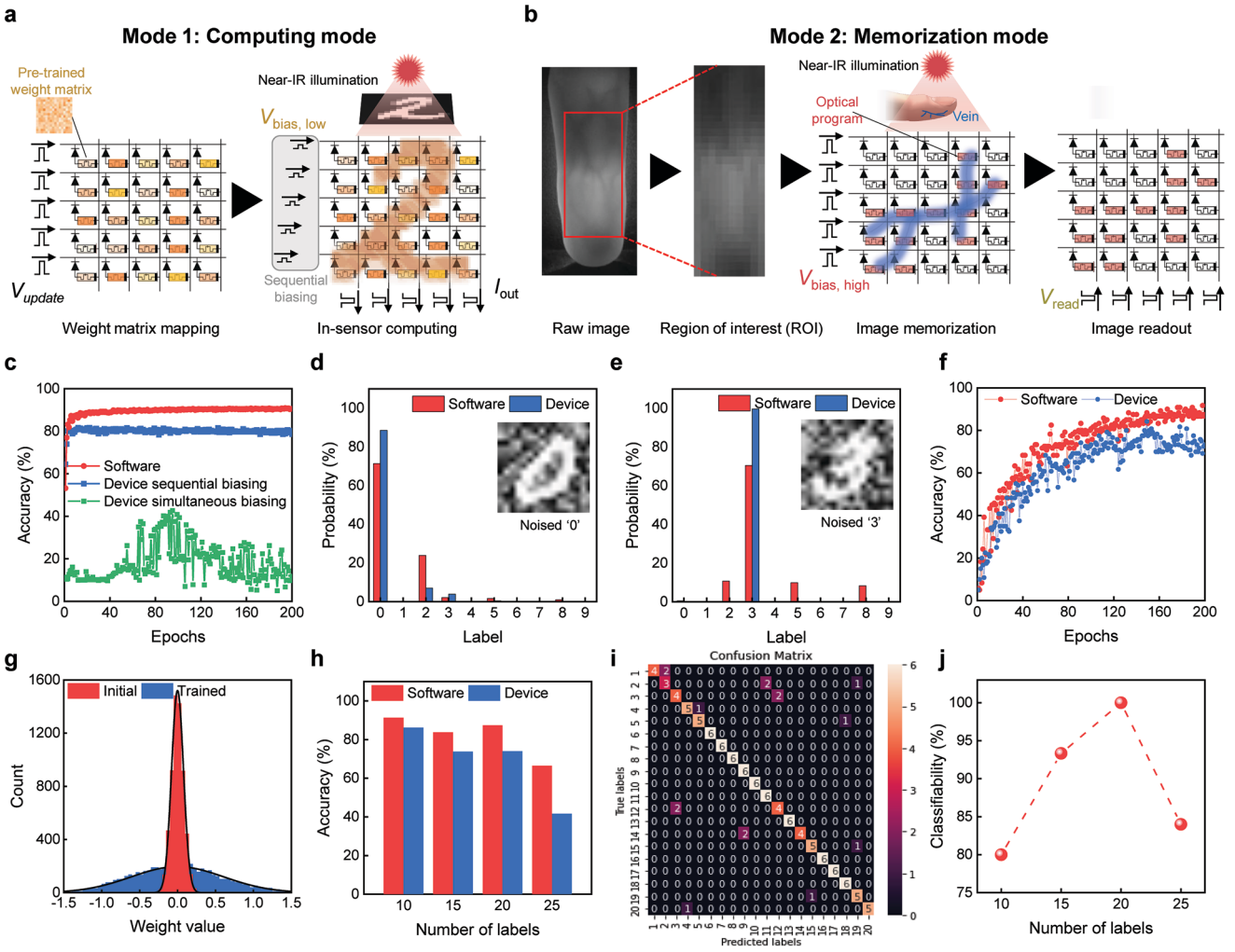

Figure 4. Applications of in-sensor computing and memorization. a) Schematic illustration of in-sensor computing mode (mode 1). The pre-trained weight matrix is mapped to the memristor array, and $V_{bias,low}$ is sequentially applied to each row. The output currents of each column are accumulated (I_{out}). b) Schematic illustration of in-sensor memorization mode. The extracted vein dataset is illuminated on the 1P-1R artificial retina with $V_{bias,high}$. c) Results of simulated classification accuracy of mode 1. The device scheme with sequential biasing exhibits classification performance (up to 82%), while the accuracy of the device scheme with conventional simultaneous is up to 43%. Inference results for noise-added d) "0" and e) "3" images. Both the software and device schemes identified the noised images correctly. f) Results of simulated classification accuracy. Up to 84% accuracy is achieved for the device scheme. g) Evolution of weight values in the hidden layer. The initial weight values are updated with the training epochs. h) Accuracy of various numbers of labels, averaged from the 151st to the 200th epoch. i) Confusion matrix of 20 labels. j) Classifiability of various numbers of labels computed from the confusion matrix. For 20 labels, the highest classifiability is achieved (100%).

We have counted the number of the "classifiable" labels, which corresponds to the numerator of the defined classifiability. By comparing the classifiability of the various numbers of labels, the 20 labels perform the best classifiability (100%), implying that most of the test data for all labels are classifiable as shown in Figure 4j (for more details, see Experimental Section).

3. Conclusion

In this study, we have fabricated the 1P-1R artificial retina capable of NIR in-sensor computing. We have employed oxygen vacancy-based NVMs to enable neuromorphic vision computing in conjunction with the InGaAs photodiode that covers the NIR spectrum. The received 2D optical input is directly recorded or computed at each pixel by coupling the resistive switching of the 1R cell with the generated photocurrent of the 1P cell. The 1P-1R artificial optical nerve offers two reconfigurable in-sensor protocols: the in-sensor computing mode and the in-sensor memorization mode. We have demonstrated the proof-of-concept applications of each mode, including MNIST and finger vein classifications. We believe the proposed 1P-1R artificial retina can provide a pathway toward neuromorphic vision computing based on the in-sensor computing/memorization functions, which are applicable to various relevant field including computer vision, biomedical engineering, artificial intelligence on things (AIoTs), and robotics.

4. Experimental Section

Device Fabrication: On an InP substrate, InGaAs p-i-n layers were fabricated using general molecular beam epitaxy (MBE).[19] The p-InGaAs/Uid-InGaAs area was etched with inductively coupled plasma-reactive ion etching (ICP-RIE) for 6 min under the conditions of BCl_3 20 sccm, 600 W ICP power, 150 W forward power, 7 mTorr, and 20 °C stage temperature, followed by a wet etching process. n-InP mesa for the bottom metal electrodes was patterned via a half-minute wet etch operation using a solution of $HCl:H_2PO_4$ = 3:1. Next, a 150 nm Al_2O_3 dielectric insulator layer was formed via plasma-enhanced atomic layer deposition (PEALD). With a bilayer PR mask and ICP-RIE under the conditions of BCl_3: 20 sccm, 50 W ICP power, 200 W forward power, 5 mTorr, and 20 °C stage temperature for 6 min, the via holes were etched. The top and bottom electrodes of the photodiodes were deposited simultaneously using photolithography with bilayer PR and e-beam evaporation of Ti/Pt/Au (5/10/50 nm). PEALD was used to deposit an additional dielectric insulator of 150 nm Al_2O_3, and through holes were made in the bottom electrodes of the photodiodes. The bottom electrodes of the NVMs, which are coupled to the bottom electrodes of the photodiodes, were subsequently deposited via e-beam evaporation of Ti/Pt (5/25 nm). On the top electrodes of the NVM, 5.5-nm HfO_2 was deposited using PEALD, followed by metal deposition of Ta/Pt (50/25 nm) through DC magnetron sputtering of Ta (25 W RF power, 5 mTorr, Ar 20 sccm, room temperature for 18 min) and e-beam evaporation of Pt.

Electrical Characterizations: Electrical measurements were performed using a semiconductor parameter analyzer (B1500A, Keysight), pulse generator (33600A, Keysight), and an oscilloscope (DSO-X 3024T, Keysight). A NIR LED (Marktech Optoelectronics / MTE2077N1-R) was used for light illumination with designed circuit using breadboard power supply (BBP-32701). The light intensity of the LED was controlled by a connected resistor and the intensity of light was measured with a USB power meter with standard photodiode sensors (PM16-120, Thorlabs).

MNIST Classification Simulations: The MNIST classification task based on the NeuroSim simulator was simulated with experimentally extracted memristor parameters.[31] The number of training and test images were 2500 and 10 000, respectively. To match the pixel dimensions with the fabricated 1P-1R artificial retina, the original 28×28 MNIST images were resized to 16×16. The ANN architecture included the in-sensor computing layer and 256-64-32-10 fully connected layers. The initial learning rate was 0.12, and the momentum was 0.7. All the weight values in each ANN layer were initialized with a Gaussian distribution, and the feed forwarding batch size was 32 with 200 epochs.

FV-USM Classification Simulations: The FV-USM classification task based on the NeuroSim simulator was simulated with experimentally extracted memristor parameters.[31] The numbers of training and test images were 18 and 6 per each label, respectively. The ROI images were resized to 30×10. The readout weight values were flattened and applied to a single fully connected layer ($300–n_{label}$, where n_{label} is the total number of the labels). The constant learning rate was 0.02 with a stochastic gradient descendent algorithm. All the weight values in each ANN layer were initialized with a Gaussian distribution.

Acknowledgements

B.B., M.P., and D.L. contributed equally to this study. This work was supported by the National Science Foundation (NSF) (Grant No. 1942868).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

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

Keywords

artificial neural networks, edge-computing, image classification, in-sensor computing, neuromorphic computing

Received: August 15, 2022
Revised: October 24, 2022
Published online: December 8, 2022

[1] L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, *Nature* **2020**, *579*, 62.
[2] S. Goossens, G. Navickaite, C. Monasterio, S. Gupta, J. J. Piqueras, R. Pérez, G. Burwell, I. Nikitskiy, T. Lasanta, T. Galán, E. Puma, A. Centeno, A. Pesquera, A. Zurutuza, G. Konstantatos, F. Koppens, *Nat. Photonics* **2017**, *11*, 366.
[3] M. S. Kim, G. J. Lee, C. Choi, M. S. Kim, M. Lee, S. Liu, K. W. Cho, H. M. Kim, H. Cho, M. K. Choi, N. Lu, Y. M. Song, D. H. Kim, *Nat. Electron.* **2020**, *3*, 546.
[4] T. Phong Truong, H. Toan Le, T. Thi Nguyen, *J. Phys. Conf. Ser.* **2020**, *1432*, 012068.
[5] F. Zhou, Y. Chai, *Nat. Electron.* **2020**, *3*, 664.
[6] A. Bhargav, P. Huynh, *Sensors* **2021**, *21*, 8203.
[7] H. S. Lee, Y. Baek, Q. Lin, J. Minsu Chen, M. Park, D. Lee, S. Kim, K. Lee, *Adv. Intell. Syst.* **2021**, *3*, 2000202.
[8] H. Kim, M. R. Mahmoodi, H. Nili, D. B. Strukov, *Nat. Commun.* **2021**, *12*, 5198.
[9] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, *Nature* **2020**, *577*, 641.
[10] H. Ryu, H. Wu, F. Rao, W. Zhu, *Sci. Rep.* **2019**, *9*, 20383.
[11] J.-H. Ryu, C. Mahata, S. Kim, *J. Alloys Compd.* **2021**, *850*, 156675.
[12] Z. Wang, H. Wu, G. W. Burr, C. S. Hwang, K. L. Wang, Q. Xia, J. J. Yang, *Nat. Rev. Mater.* **2020**, *5*, 173.
[13] Q.-B. Zhu, B. Li, D.-D. Yang, C. Liu, S. Feng, M.-L. Chen, Y. Sun, Y.-N. Tian, X. Su, X.-M. Wang, S. Qiu, Q.-W. Li, X.-M. Li, H.-B. Zeng, H.-M. Cheng, D.-M. Sun, *Nat. Commun.* **2021**, *12*, 1798.
[14] M. Park, Y. Baek, M. Dinare, D. Lee, K.-H. Park, J. Ahn, D. Kim, J. Medina, W.-J. Choi, S. Kim, C. Zhou, J. Heo, K. Lee, *Sci. Rep.* **2020**, *10*, 2764.
[15] Y. Luo, J. Remillard, D. Hoetzer, in *2010 IEEE Intelligent Vehicles Symp.*, IEEE, Piscataway, NJ **2010**, pp. 51–58.
[16] A. King, *Nature* **2017**, *544*, S21.
[17] C. Guan, L. G. Hassebrook, D. L. Lau, V. Yalla, *J. Visual Commun. Image Represent.* **2007**, *18*, 141.
[18] C.-T. Pan, M. D. Francisco, C.-K. Yen, S.-Y. Wang, Y.-L. Shiue, *Sensors* **2019**, *19*, 3573.
[19] D. Fan, K. Lee, S. R. Forrest, *ACS Photonics* **2016**, *3*, 670.
[20] L. D. Moldavskaya, V. I. Shashkin, M. N. Drozdov, Y. N. Drozdov, V. M. Daniltsev, A. V. Murel, B. A. Andreev, A. N. Yablonsky, S. A. Gusev, D. M. Gaponova, O. I. Khrykin, A. Y. Luk'yanov, E. N. Sadova, *Phys. E Low Dimens. Syst. Nanostruct.* **2003**, *17*, 634.
[21] X. Lu, J. Vaillancourt, M. J. Meisner, *Semicond. Sci. Technol.* **2007**, *22*, 993.
[22] T. Maruyama, F. Narusawa, M. Kudo, M. Tanaka, Y. Saito, A. Nomura, *Opt. Eng.* **2002**, *41*, 395.
[23] M. Park, Y. Yuan, Y. Baek, A. H. Jones, N. Lin, D. Lee, H. S. Lee, S. Kim, J. C. Campbell, K. Lee, *Adv. Intell. Syst.* **2022**, *4*, 2100159.
[24] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, *Proc. IEEE* **1998**, *86*, 2278.
[25] M. S. Mohd Asaari, S. A. Suandi, B. A. Rosdi, *Expert Syst. Appl.* **2014**, *41*, 3367.
[26] P. M. Iuvone, in *Encyclopedia of the Eye* (Ed: D. A. Dartt), Academic Press, Oxford **2010**, p. 130.

[27] R. H. Masland, *Neuron* **2012**, *76*, 266.
[28] J. B. Demb, J. H. Singer, *Annu. Rev. Vision Sci.* **2015**, *1*, 263.
[29] M. Wilson, D. I. Vaney, in *The Senses: A Comprehensive Reference* (Eds: R. H. Masland, T. D. Albright, T. D. Albright, R. H. Masland, P. Dallos, D. Oertel, S. Firestein, G. K. Beauchamp, M. Catherine Bushnell, A. I. Basbaum, J. H. Kaas, E. P. Gardner), Academic Press, New York **2008**, p. 361.
[30] S. A. Bloomfield, in *Encyclopedia Of Neuroscience* (Ed: L.R. Squire), Academic Press, Oxford, **2009**, p. 171.
[31] P. Chen, X. Peng, S. Yu, *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **2018**, *37*, 3067.