

# Optoelectronic graded neurons for bioinspired in-sensor motion perception

Received: 29 November 2022

Accepted: 17 March 2023

Published online: 20 April 2023



Jiewei Chen<sup>1,2,5</sup>, Zheng Zhou<sup>3,5</sup>, Beom Jin Kim<sup>4</sup>, Yue Zhou<sup>1</sup>, Zhaoqing Wang<sup>1</sup>, Tianqing Wan<sup>1</sup>, Jianmin Yan<sup>1</sup>, Jinfeng Kang<sup>3</sup>, Jong-Hyun Ahn<sup>4</sup>✉ & Yang Chai<sup>1,2</sup>✉

Motion processing has proven to be a computational challenge and demands considerable computational resources. Contrast this with the fact that flying insects can agilely perceive real-world motion with their tiny vision system. Here we show that phototransistor arrays can directly perceive different types of motion at sensory terminals, emulating the non-spiking graded neurons of insect vision systems. The charge dynamics of the shallow trapping centres in MoS<sub>2</sub> phototransistors mimic the characteristics of graded neurons, showing an information transmission rate of 1,200 bit s<sup>-1</sup> and effectively encoding temporal light information. We used a 20 × 20 photosensor array to detect trajectories in the visual field, allowing the efficient perception of the direction and vision saliency of moving objects and achieving 99.2% recognition accuracy with a four-layer neural network. By modulating the charge dynamics of the shallow trapping centres of MoS<sub>2</sub>, the sensor array can recognize motion with a temporal resolution ranging from 10<sup>1</sup> to 10<sup>6</sup> ms.

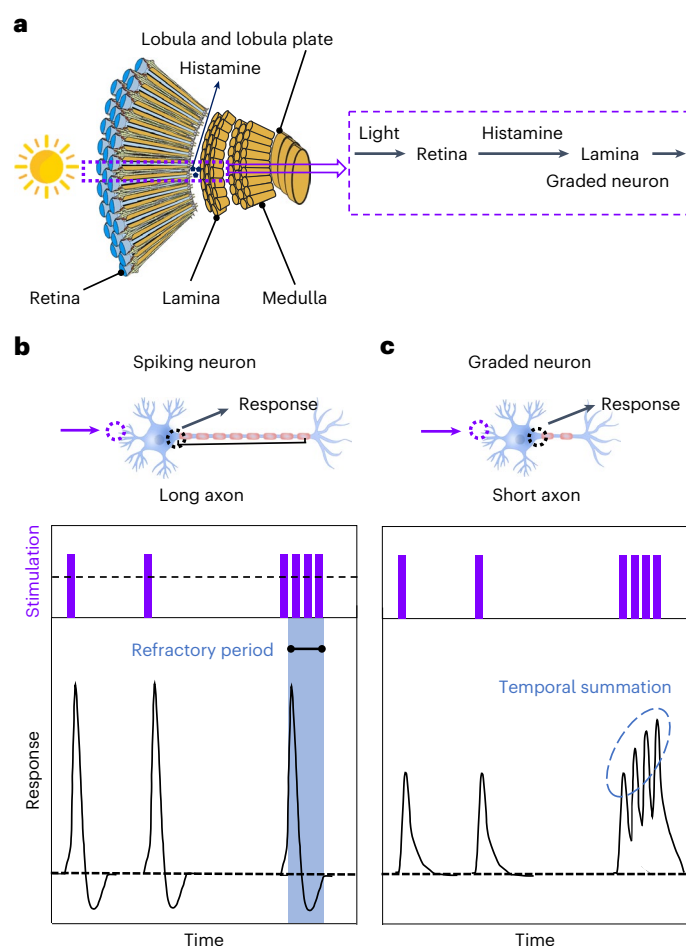
Visual information makes up a large proportion of the massive data held in the Internet of Things<sup>1–4</sup>. Ubiquitously distributed image sensors are required to recognize static images<sup>5,6</sup> and dynamic motions<sup>7,8</sup> in power-constraint scenarios and acquire understanding in an intelligent way<sup>9</sup>, playing critical roles in a range of applications, including automatic vehicles and surveillance systems. State-of-the-art machine vision systems usually consist of hardware with physically separated image sensors and processing units<sup>9,10</sup>. Most image sensors can only output spatial frames without fusing temporal information<sup>11,12</sup>. For accurate motion recognition, it is necessary for ‘spatial’ and ‘temporal’ stream information to be transferred to and fused in the processing units. Therefore, machine vision systems for action recognition typically involve complex artificial neural networks, for example, ‘spatial’ and ‘temporal’ stream computation architectures with ~2 × 8 layers and ~10<sup>6</sup> network parameters (Supplementary Fig. 1 and Supplementary Note 1)<sup>13,14</sup>.

Biological visual systems can effectively perceive motion in a complicated environment with high energy efficiency<sup>15–18</sup>. Flying insects

such as *Drosophila* with a tiny vision system (~8 × 10<sup>2</sup> ommatidia and 10<sup>5</sup> neurons<sup>19,20</sup>) can agilely recognize a moving object much faster than humans. The insect visual system consists of non-spiking graded neurons (retina–lamina) that have a much higher information transmission rate (*R*) than the spiking neurons in the human visual system. More importantly, the graded neurons enable efficient encoding of temporal information at sensory terminals, which reduces the transfer of abundant vision data for fusing spatiotemporal information in a computation unit<sup>21–23</sup>. This tiny and acute insect visual system can effectively perceive motion with limited computational resources, which has inspired us to develop optoelectronic graded neurons for in-sensor motion perception.

In this work we experimentally investigated optoelectronic graded neurons for agile in-sensor motion perception. The individual bioinspired vision sensor can effectively respond and encode light stimuli in the temporal domain by emulating the graded neurons of the insect visual system. We used two-dimensional (2D) atomically thin semiconductors to emulate the optoelectronic graded neuron

<sup>1</sup>Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong, China. <sup>2</sup>Shenzhen Research Institute, The Hong Kong Polytechnic University, Shenzhen, China. <sup>3</sup>School of Integrated Circuits, Peking University, Beijing, China. <sup>4</sup>School of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea. <sup>5</sup>These authors contributed equally: Jiewei Chen, Zheng Zhou. ✉e-mail: [ahnj@yonsei.ac.kr](mailto:ahnj@yonsei.ac.kr); [ychai@polyu.edu.hk](mailto:ychai@polyu.edu.hk)



**Fig. 1 | Agile motion perception of the insect visual system.** **a**, Schematic of the tiny insect visual system. Light stimulation can cause a graded membrane potential in photoreceptors in the retina and the release of histamine, which transfers to the lamina neuron (a typical graded neuron) for temporal processing. The lamina neurons transmit the processed signals to the medulla and lobula layers for action recognition. **b**, The structure (top) and response characteristics (bottom) of a spiking neuron. Strong stimulation can trigger the action potential, and then the neuron enters the refractory period. The spiking neuron does not respond to stimulation during the refractory period. **c**, The structure (top) and response characteristics (bottom) of a graded neuron. A graded neuron can respond to sequential stimulation with nonlinear temporal summation characteristics.

array that can encode spatiotemporal motion information at sensory terminals. The in-sensor fused spatiotemporal motion information is fed into a tiny four-layer perceptron that exhibits a high accuracy (99.2%) for moving-ball track recognition, which is scarcely achievable in conventional image sensors with such a tiny network. In addition, the  $20 \times 20$  sensor array can effectively perceive motion of different speeds by modulating the charge-trapping dynamics of artificially graded neurons.

## Agile motion perception of the insect visual system

Flying insects can acutely perceive motion, which allows them to respond agilely in a cluttered and uncertain environment (Supplementary Fig. 2). The agility of the visual system can be quantitatively described by the flicker fusion frequency (FFF). The FFF of *Drosophila* is  $\sim 100$  Hz (ref. 24), four times faster than that of humans<sup>25</sup>. There are two reasons for the agile motion perception of insects. (1) Their tiny vision

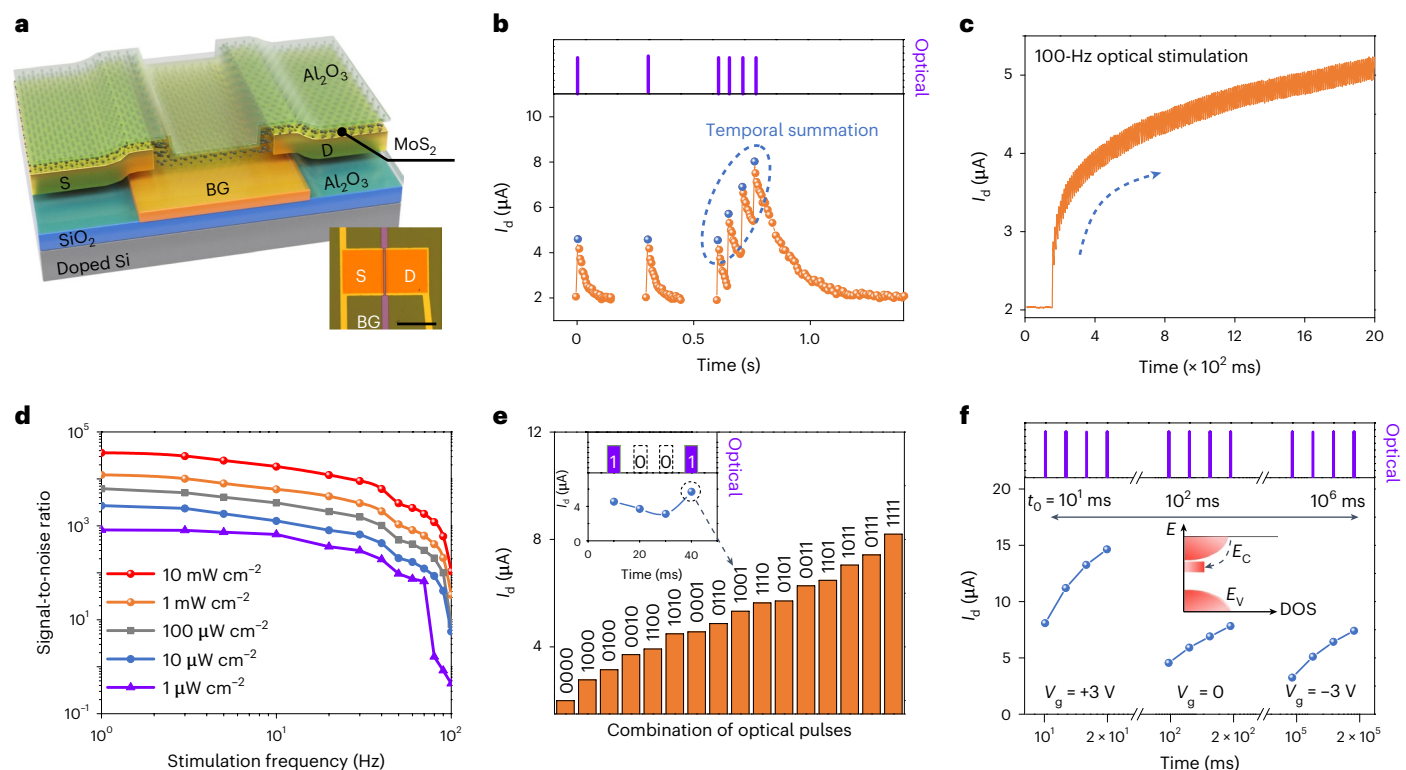
system greatly decreases the signal transmission distance between the retina (sensor) and brain (computation unit). Figure 1a schematically illustrates the tiny vision system of insects, which includes the retina (photoreceptors in the ommatidia), lamina, medulla and lobula. (2) The lamina neurons can effectively encode temporal information after receiving histamine from the photoreceptor<sup>26–28</sup>. The non-spiking graded signal transmission rate between the retina and lamina neurons reaches  $1,650 \text{ bit s}^{-1}$ , approximately five times higher than that in spiking neurons ( $\sim 300 \text{ bit s}^{-1}$ )<sup>21,29</sup>. The existence of a refractory period in spiking neurons greatly limits the speed of information transmission. The volatile characteristics of spiking and graded neurons<sup>30</sup> are compared in Fig. 1b,c and Supplementary Table 1. Supplementary Fig. 3 shows the temporal response of spiking and graded neurons (Supplementary Note II). Unlike spiking neurons, which exhibit ‘all-or-none’ characteristics, graded neurons (for example, lamina) show multi-level responses instead of the spiking feature. The graded response depends on the strength of sequential stimulations, with graded neurons exhibiting temporal summation characteristics without a refractory period. The non-spiking graded characteristics also enable lamina neurons to efficiently encode the spatiotemporal information of moving objects at sensory terminals (Supplementary Fig. 4)<sup>31</sup>.

## Artificially graded neurons for encoding temporal vision

Two-dimensional phototransistors provide a hardware platform for emulating the functionalities of graded neurons (for example, temporal summation) owing to their shallow charge-trapping centres. We fabricated bilayer  $\text{MoS}_2$  phototransistors (Fig. 2a) that showed good uniformity (Supplementary Fig. 5). The extracted trapped charge density of the  $\text{MoS}_2$  phototransistors was  $\sim 1 \times 10^{12} \text{ cm}^{-2}$ , as determined from the hysteresis voltage window (Supplementary Fig. 5a and Supplementary Note III), which arises from the intrinsic defects of the as-grown  $\text{MoS}_2$  film. The coupling of light stimulation and shallow trapping centres in  $\text{MoS}_2$  results in volatile characteristics (Supplementary Fig. 6). The charge-trapping mechanism was verified through temperature-dependent measurements (Supplementary Fig. 7)<sup>32</sup>. Thus, light stimuli trigger photogenerated carriers in the  $\text{MoS}_2$  film, which can mimic the phototransduction process of the retina (Supplementary Note IV and stage I in Supplementary Fig. 8). The photogenerated carriers trapped in the gap states of  $\text{MoS}_2$  result in a sublinear increase in the conductance in the temporal domain, emulating the release of histamine to the lamina neuron (stage II in Supplementary Fig. 8).

We then characterized the bilayer  $\text{MoS}_2$  phototransistors in a series of experiments. Under light illumination (5 ms pulse width with an intensity of  $5 \text{ mW cm}^{-2}$  at 660 nm), the drain current ( $I_d$ ) of the  $\text{MoS}_2$  phototransistor increased dramatically (Fig. 2b). After the light illumination,  $I_d$  decreased to zero due to the release of charges from the shallow trapping centres; the decay can be described by an exponential relationship with a time constant ( $t_0$ ; Supplementary Note V and Supplementary Fig. 9). The parameter  $t_0$  represents the time for  $I_d$  to reach  $1/e$  ( $\sim 36.8\%$ ) of the peak value following light stimulation. For zero gate voltage ( $V_g = 0$ ), the extracted  $t_0$  was  $\sim 100$  ms after a single light pulse (Fig. 2b). When the time interval between two adjacent light pulses was longer than  $t_0$  ( $\sim 100$  ms), the two responses to the light pulses were nearly independent. After four sequential stimulation pulses with a time interval between pulses shorter than  $t_0$ , the photosensor sums in a graded manner the light stimuli in the temporal domain through nonlinear accumulation, emulating the characteristics of graded neurons (Fig. 1c).

Upon light stimulation of  $5 \text{ mW cm}^{-2}$  at 660 nm with a frequency of 100 Hz,  $I_d$  increased gradually and tended to saturate (Fig. 2c). Because  $t_0 \approx 100$  ms, the charge trapped after light illumination was incompletely released during the interval between pulses (10 ms). Subsequent light illumination generated more charge and filled more trap states, resulting in a higher  $I_d$ . This nonlinear increase in  $I_d$  emulates the response of



**Fig. 2 | Artificially graded neurons for encoding temporal vision.** **a**, Schematic of the MoS<sub>2</sub> phototransistor. The inset shows an optical microscopy image of the MoS<sub>2</sub> phototransistor with source (S), drain (D) and bottom gate (BG) terminals. Scale bar, 200  $\mu\text{m}$ . **b**, Photocurrent as a function of time under single light pulse stimulation. Applied stimulation light pulse: 660 nm, 5  $\text{mW cm}^{-2}$  and 10 ms. The applied voltage between drain and source ( $V_{\text{ds}}$ ) was 0.1 V at  $V_{\text{g}} = 0$ . 'Optical' indicates the light stimulation. **c**, Photocurrent as a function of illumination time for a stimulation frequency of 100 Hz. Applied stimulation light pulse: 660 nm, 5  $\text{mW cm}^{-2}$  and  $\sim 5$  ms. **d**, Signal-to-noise ratio under different light

intensities ranging from 1  $\mu\text{W cm}^{-2}$  to 10  $\text{mW cm}^{-2}$ . The stimulation frequency was varied from 1 to 100 Hz with a duty ratio of 50%. **e**, Responses to four-bit light stimulation. The inset shows the encoding process for '1001' stimulation, where '1' and '0' represent 5  $\text{mW cm}^{-2}$  light illumination and dark conditions, respectively. **f**, Processing by MoS<sub>2</sub> phototransistors of four light stimulation pulses for time constants varying from 10<sup>1</sup> ms ( $V_{\text{g}} = +3$  V) to 10<sup>6</sup> ms ( $V_{\text{g}} = -3$  V) under different  $V_{\text{g}}$ . The inset shows the band structure of the bilayer MoS<sub>2</sub> and the trapping centre.  $E$  is the energy,  $E_{\text{c}}$  is the conduction band,  $E_{\text{v}}$  is the valence band and DOS is the density of states.

lamina neurons (Supplementary Fig. 3). The signal-to-noise ratio (SNR) is the ratio between the signal power spectra and the noise power spectra (see Supplementary Note VI for details of calculations). Sublinear response characteristics can improve the SNR, especially under dim light (Supplementary Fig. 10 and Supplementary Note VII). FFF is the critical frequency at which flickering light (generally, 50% duty ratio) is perceived as continuous ( $\text{SNR} \leq 1$ ) and has been widely used to assess temporal vision processing. The photosensor showed a distinguishable response under light stimulation with a frequency of 100 Hz (Fig. 2c), suggesting an FFF of at least 100 Hz. The SNR varied as a function of light intensity from 1  $\mu\text{W cm}^{-2}$  to 10  $\text{mW cm}^{-2}$  (Fig. 2d). The FFFs under light intensities of 10  $\mu\text{W cm}^{-2}$  and 10  $\text{mW cm}^{-2}$  were determined to be at least 70 and 100 Hz, respectively, that is, they increase with light intensity, matching very well the results of biological experiments.

$R$  characterizes the information capacity after receiving vision stimulation and can be extracted from the frequency-dependent SNR according to the Shannon equation (Supplementary Note VIII). On increasing the light stimulation frequency from 1 to 100 Hz, we extracted values of  $R$  of 1,000 and 1,200  $\text{bit s}^{-1}$  under light intensities of 1 and 10  $\text{mW cm}^{-2}$ , respectively, much higher than that of spiking neurons ( $\sim 300 \text{ bit s}^{-1}$ )<sup>21</sup>. The high  $R$  value of optoelectronic graded neurons can result in agile motion perception, while the limited spike rate of spiking neurons restricts the information transmission rate<sup>33</sup>.

We tested the photoresponse of the bioinspired sensor under a series of sequential light pulses. The inset of Fig. 2e shows a typical encoding process for '1001' with the graded changes at four points in

time. The digits '1' and '0' represent 5  $\text{mW cm}^{-2}$  light illumination and dark conditions, respectively. Figure 2e shows the exported  $I_{\text{d}}$  under four light pulses (16 combinations) with a width of 50 ms and interval of 50 ms, revealing distinctive responses to the different four-frame actions that enables encoding processes. The  $I_{\text{d}}$  is greatly dependent on the sequence of light stimulation. When the time interval between '1' stimulation pulses is much shorter than  $t_0$ , the stimulation pulses are temporally summed, leading to a relatively high output of  $I_{\text{d}}$ ; when the '1' stimulation pulses are separated in the temporal domain (similar to or longer than  $t_0$ ),  $I_{\text{d}}$  is relatively low due to the release of trapped electrons. The distinguishable outputs suggest that the photosensor can temporally encode four-pulse light stimulation similarly to graded neurons.

The speeds of different motions vary vastly. Oversampling results in data redundancy, while undersampling misses motion information. Motion perception with appropriate temporal resolution can achieve a balance between efficiency and accuracy.  $I_{\text{d}}$  shows a fast/slow decay time due to shallow/deep trap states<sup>34</sup>. The  $V_{\text{g}}$  of phototransistors allows tuning of the trapping/detrapping characteristics and decay curves of  $I_{\text{d}}$ , which can modulate  $t_0$  from 10<sup>1</sup> to 10<sup>6</sup> ms (Supplementary Fig. 11 and Supplementary Note IX). Upon light stimulation with gate voltages of  $-3 \text{ V} \leq V_{\text{g}} \leq +3 \text{ V}$ , shallow trapping centres, which show graded neuron characteristics instead of synaptic plasticity features, are dominant. Under a negative  $V_{\text{g}}$ , photogenerated electrons are trapped at a relatively deep level; a positive  $V_{\text{g}}$  inhibits the trapping of photogenerated electrons<sup>15</sup>. Figure 2f shows the  $I_{\text{d}}$  of the phototransistor under different



$V_g$  (+3, 0 and -3 V). The sampling frequency was set to  $1/t_0$  ( $t_0$  values of 10,  $10^2$  and  $10^6$  ms correspond to sampling frequencies of  $10^2$ ,  $10^1$  and  $10^{-3}$  Hz, respectively; Supplementary Fig. 12). Under  $V_g = +3$  V,  $t_0$  is -10 ms, which allows fast motions to be processed. Under  $V_g = -3$  V, the MoS<sub>2</sub> phototransistor has a  $t_0$  value of - $10^6$  ms, which is appropriate for processing slow motions. This ability to tune the  $t_0$  of phototransistors enables dynamic vision tasks to be processed with a temporal resolution ranging from  $10^1$  to  $10^6$  ms.

## In-sensor motion perception with a graded neuron array

The time-dependent characteristics of the MoS<sub>2</sub> photosensor enable it to encode sequentially temporal information. We used a photosensor array ( $20 \times 20$  pixels) that can encode spatiotemporal vision information and show the contour of the trajectory in the visual field, thus allowing dynamic motion to be perceived (Fig. 3a). The design of the photosensor array for in-sensor motion perception is shown in Supplementary Fig. 13 and Supplementary Note X. Each device (optoelectronic graded neuron) with a local bottom-gate structure can individually process pixel-level temporal sequences. We used four temporal frames to represent specific actions. Supplementary Fig. 14 shows the temporal evolution of the visual stimulation of four motions (moving right, moving left, approaching and leaving) projected onto a  $20 \times 20$  sensor array. A high SNR is critical for accurate and fast motion perception. The SNR under all the tested frequencies was higher than 1,000 (Supplementary Fig. 15), indicating that the devices can robustly perceive motion with different temporal resolution.

The perception of motion direction is important in visual processing. We used a sampling frequency of 10 Hz with  $V_g = 0$  to process four temporal frames (Supplementary Fig. 14a,b). As the motion proceeded, the photocurrents gradually decayed with elapsed time after light stimulation. The next frame showed a higher photocurrent than the previous frame. The superposition of the frames obtained with the bioinspired sensor array can consolidate the spatiotemporal motion information through the distinctive level of the photocurrent, clearly showing the contour of the trajectory of the entire motion from left to right (Fig. 3b(i) and Supplementary Fig. 16a). When the motion was in the reverse direction, the sensor array could perceive the reversed contour of the trajectory (Fig. 3b(ii) and Supplementary Fig. 16b), emulating the output of lamina neuron arrays. In contrast, conventional complementary metal-oxide semiconductor image sensors can only output the latest frame of the motion, thereby lacking the ability to encode temporal vision and perceive motion.

The characteristics of the artificially graded neurons also allow the perception of approaching and leaving motion. When we reduced the distance between the sensor and object, the projected size increased in the visual field (Supplementary Fig. 14c,d). The photocurrent contour maps exhibited an expanding image, indicating approaching motion (Fig. 3c(i) and Supplementary Fig. 16c). When the distance between the sensor and object was increased, the photocurrent contour displayed a contracting image (Fig. 3c(ii) and Supplementary Fig. 16d). These distinctive photocurrent contours of motion make it easy to judge approaching and leaving motion.

When the speed of right-to-left movement was increased from  $0.5$  to  $5 \text{ m s}^{-1}$ , the object quickly moved outside the visual field. Under  $V_g = 0$ , the long  $t_0$  (100 ms) allowed only the contour of the initial frame to be recorded (Fig. 3d(i)). To enable the perception of movement with

different motion speeds, we modulated  $V_g$  from 0 to +3 V. The value of  $t_0$  under  $V_g = +3$  V is 10 ms, an order of magnitude lower than that under  $V_g = 0$ , which enables motion with a speed of  $5 \text{ m s}^{-1}$  to be traced (Fig. 3d(ii)). Thus, the bioinspired sensor arrays can perceive motion with different speeds by modulating  $V_g$ .

Insect visual systems can track objects of interest (target salience) and neglect objects that move too slowly or too quickly. Flying insects have high FFF and tend to perceive objects with high movement speeds. We tested the target salience of our device by introducing both slow ( $1 \text{ m s}^{-1}$ ) and fast ( $10 \text{ m s}^{-1}$ ) motion into the same visual scene. The sensor array was able to record the trajectory of the fast-moving object under  $V_g = +3$  V (Supplementary Fig. 17), but under the same testing conditions, the trajectory contour of the slow motion could not be detected. Therefore, the sensor array exhibits target salience characteristics. This vision salience enables bioinspired vision sensors to focus on the targeted motion and filter the unintentional one, thereby greatly reducing the requirement for subsequent computation resources.

## Highly accurate motion recognition

High-accuracy agile motion recognition is essential for machine vision applications (for example, automated vehicles and surveillance). We used bioinspired in-sensor motion perception to classify the motion directions of a ball (Fig. 4a). We developed a customized dataset including five types of real-world motion (moving up, down, left, right and blank; typical actions are shown in Supplementary Video 1) for vision system training and recognition. The sensor array can feed these captured images into a tiny artificial neural network for recognizing actions. The bioinspired vision sensors can fuse the spatiotemporal information of a series of frames into compressive images, similar to the retina-lamina functions in *Drosophila*. In contrast, conventional image sensors output frames with only the spatial information.

Figure 4b(i) shows the five representative motions of a ball (up, down, left, right and blank). The bioinspired vision sensor output the frame with compressive temporal states (Fig. 4b(ii)), while conventional image sensors could sense only the present stimulation (Fig. 4b(iii)). The compressive frames clearly show the contour of the trajectory, which allows easy motion recognition. Figure 4c shows the recognition accuracy of the artificial neural networks with bioinspired vision sensors and conventional image sensors during the training process. The accuracy of action recognition with bioinspired vision sensors reached 99.2% after about ten training rounds, much higher than that achieved with conventional image sensors (~50%). The high recognition accuracy with the bioinspired vision sensors arises from the fused spatiotemporal information of the entire action. We investigated the effects of device-to-device photoresponse variation on the accuracy of motion perception. The recognition accuracy was higher than 90% even for a device-to-device variation of 25% (Supplementary Fig. 18), suggesting that the in-sensor motion perception tolerates a certain level of device-to-device variation.

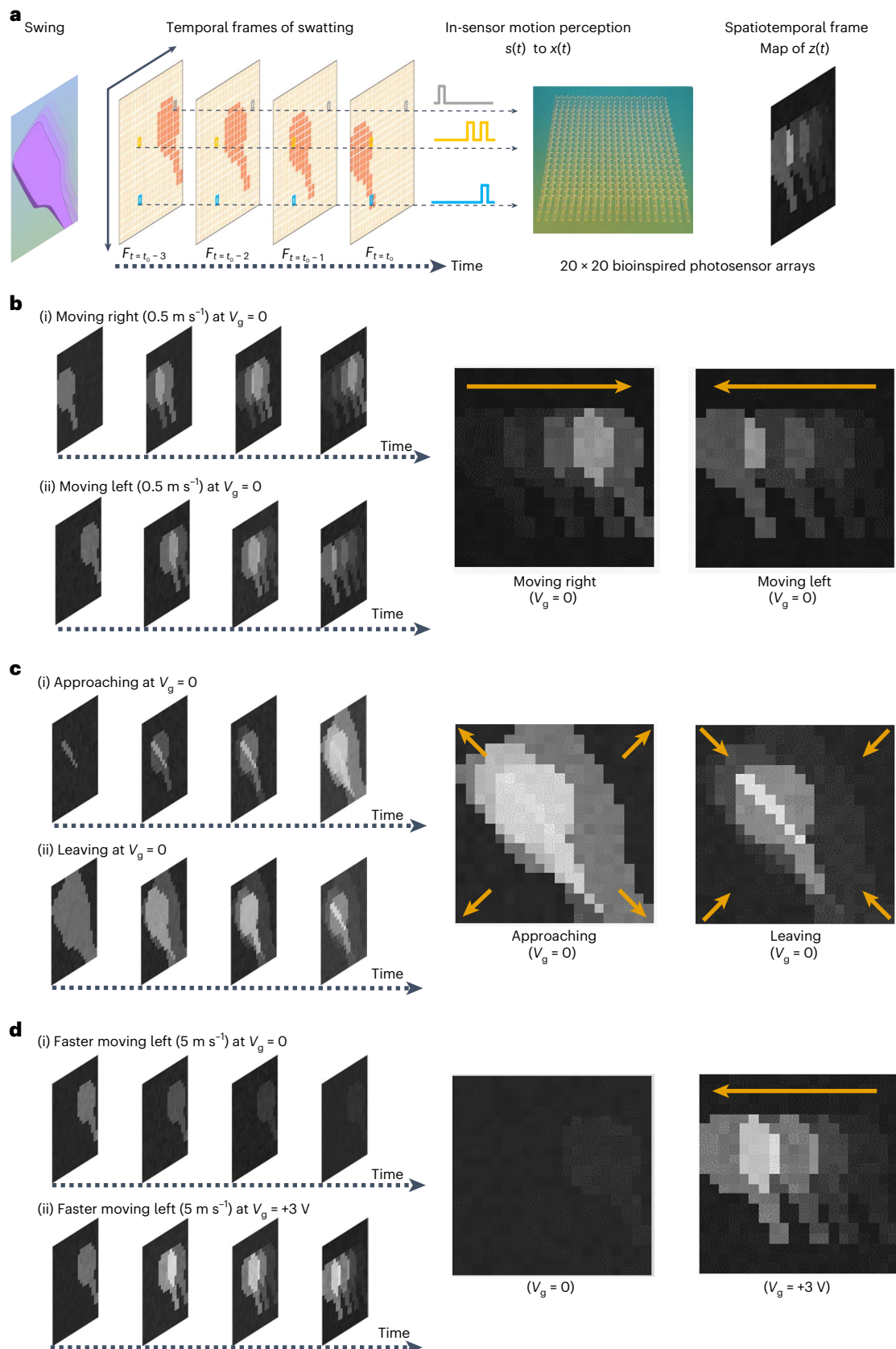
Supplementary Fig. 19 shows a typical time-related action recognition of a left-moving ball. This motion consisted of three parts: appearing, moving and disappearing within 7 s (top panel of Supplementary Fig. 19 and Supplementary Note XI). The trained bioinspired vision system accurately recognized the left-moving action, while the trained neural network based on conventional image sensors failed to recognize the motion (bottom panel of Supplementary Fig. 19 and Supplementary Video 2).

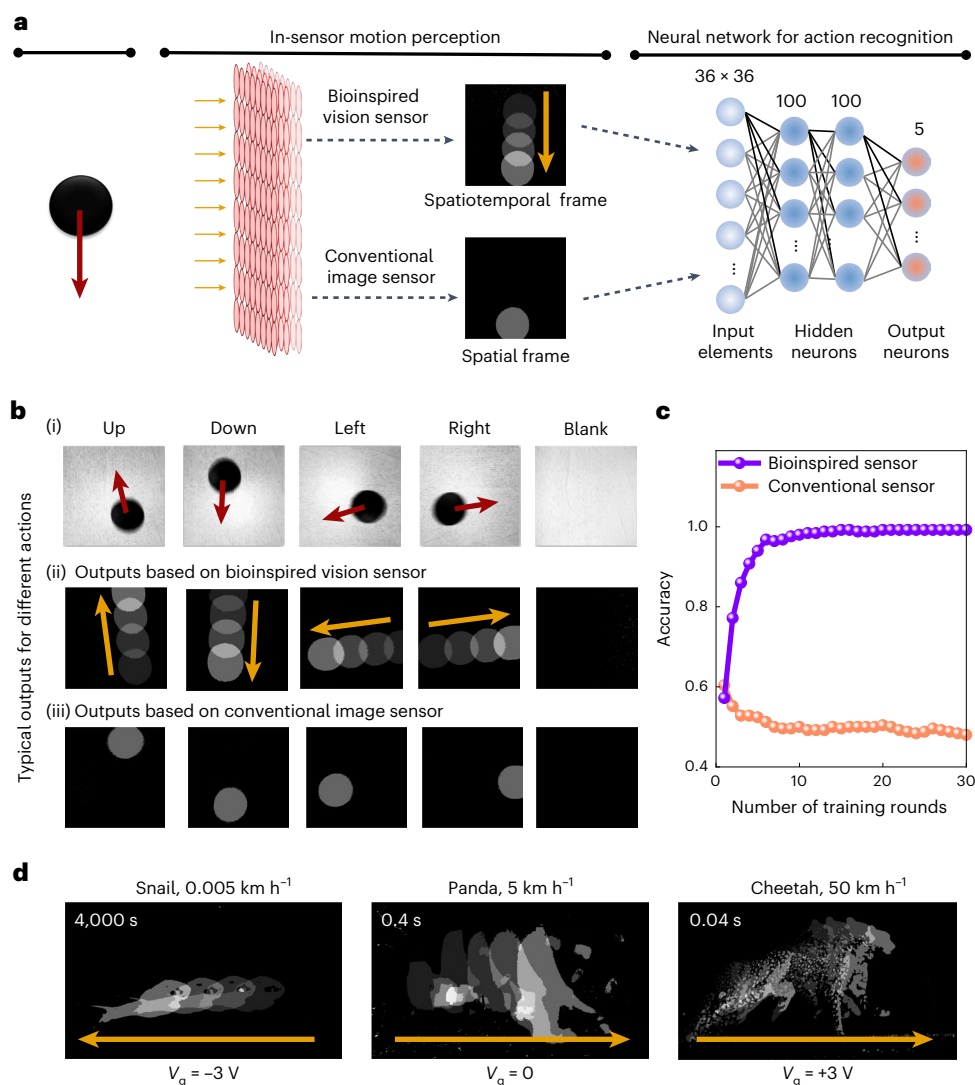
**Fig. 3 | In-sensor motion perception with an optoelectronic graded neuron array.** **a**, MoS<sub>2</sub> phototransistor arrays for mapping spatiotemporal vision information. The motion is visualized in a sequence of four frames ( $F_{t_0-3}$ ,  $F_{t_0-2}$ ,  $F_{t_0-1}$  and  $F_{t_0}$ ). The pixels in a specific column of the temporal frames form the temporal vision sequence  $s(t)$ . Bioinspired vision sensors can sense and process the temporal vision sequence  $s(t)$  into temporal compressive states  $x(t)$ . **b**, The output current maps for left-to-right (i) and right-to-left (ii) motion.

**c**, The output current maps for approaching (i) and leaving (ii) motion. The approaching motion results in a large, clear contour, while the contour of the leaving object is small and blurred. **d**, The output current maps for faster motion with a speed of  $5 \text{ m s}^{-1}$  for  $V_g = 0$  (i) and  $V_g = +3$  V (ii). The optoelectronic graded neuron array can perceive the fast-moving scene by increasing  $V_g$  from 0 to +3 V. In **b–d**, enlargements of the final photocurrent contours are shown on the right.

The speed of motion in nature varies considerably. The snail moves slowly at a speed of  $0.005 \text{ km h}^{-1}$ , the panda walks at  $5 \text{ km h}^{-1}$  and the cheetah runs as quickly as  $50 \text{ km h}^{-1}$ . Motion perception at different sampling frequencies enables a trade-off between recognition accuracy and computing resources. Adjustable temporal

summation at sensory terminals enables action recognition to be processed for most real-world scenes. Using our bioinspired vision system, we studied three left-moving balls moving at different speeds, changing the time constant of the phototransistors by modulating  $V_g$ , which allowed accurate recognition of a moving ball with different





**Fig. 4 | Action recognition based on bioinspired vision sensors and conventional image sensors.** **a**, A tiny neural network for action recognition with bioinspired vision and conventional image sensors. Bioinspired vision sensors can output spatiotemporal frames, while conventional vision sensors output only spatial frames. **b**, Typical up, down, left, right and blank motions (i), the outputs from the bioinspired sensor (ii) and the outputs from a conventional

image sensor (iii). The diameter of the moving ball was 5 cm. **c**, The action recognition accuracy of bioinspired and conventional sensors. **d**, The bioinspired photosensors can encode spatiotemporal information at different time scales, covering the motion of snails, pandas and cheetahs with distinctly different moving speeds ranging from 0.005 to 50 km h<sup>-1</sup>.

speeds (Supplementary Fig. 20). This feature enables the encoding of actions on different timescales into spatiotemporal frames (Fig. 4d) under different  $V_g$ . Thus, a negative  $V_g$  (-3 V) can process the slow movement of a snail, while a positive  $V_g$  (+3 V) can deal with the fast movement of a cheetah.

## Conclusions

The charge dynamics of the shallow trapping centres of MoS<sub>2</sub> phototransistors allow hardware devices that exhibit sublinear optoelectronic response and high information transmission rate to emulate the characteristics of non-spiking graded neurons. The artificially graded neurons enable direct sensing and encoding of the temporal information at the single-pixel level. The bioinspired vision sensor array can encode spatiotemporal visual information and display the contour of the trajectory, enabling the perception of motion with limited hardware resources. By encoding the spatiotemporal information and feeding the compressive images into an artificial neural network, the accuracy of action recognition reaches 99.2%, much higher than the recognition

accuracy achieved with conventional image sensors (~50%). Thus, the bioinspired vision sensor provides in-sensor motion perception for acute action recognition with limited computational resources.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-023-01379-2>.

## References

1. Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
2. Jang, H. et al. In-sensor optoelectronic computing using electrostatically doped silicon. *Nat. Electron.* **5**, 519–525 (2022).
3. Chai, Y. In-sensor computing for machine vision. *Nature* **579**, 32–33 (2020).

4. Choi, C. et al. Curved neuromorphic image sensor array using a MoS<sub>2</sub>-organic heterostructure inspired by the human visual recognition system. *Nat. Commun.* **11**, 5934 (2020).
5. Zhou, F. et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).
6. Seung, H. et al. Integration of synaptic phototransistors and quantum dot light-emitting diodes for visualization and recognition of UV patterns. *Sci. Adv.* **8**, eabq3101 (2022).
7. Jayachandran, D. et al. A low-power biomimetic collision detector based on an in-memory molybdenum disulfide photodetector. *Nat. Electron.* **3**, 646–655 (2020).
8. Chai, Y. Silicon photodiodes that multiply. *Nat. Electron.* **5**, 483–484 (2022).
9. Zhou, F. & Chai, Y. Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020).
10. Li, X. et al. Power-efficient neural network with artificial dendrites. *Nat. Nanotechnol.* **15**, 776–782 (2020).
11. Wan, T. et al. In-sensor computing: materials, devices, and integration technologies. *Adv. Mater.* **9**, 2203830 (2022).
12. Kim, M. et al. An aquatic-vision-inspired camera based on a monocentric lens and a silicon nanorod photodiode array. *Nat. Electron.* **3**, 546–553 (2020).
13. Simonyan, K. & Zisserman, A. Two-stream convolutional networks for action recognition in videos. *Adv. Neural Inf. Process Syst.* **27**, 568–576 (2014).
14. Ye, H. et al. Evaluating two-stream CNN for video classification. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval* 435–442 (Association for Computing Machinery, 2015).
15. Liao, F. et al. Bioinspired in-sensor visual adaptation for accurate perception. *Nat. Electron.* **5**, 84–91 (2022).
16. Jung, D. et al. Highly conductive and elastic nanomembrane for skin electronics. *Science* **373**, 1022–1026 (2021).
17. Song, Y. M. et al. Digital cameras with designs inspired by the arthropod eye. *Nature* **497**, 95–99 (2013).
18. Lee, M. et al. An amphibious artificial vision system with a panoramic visual field. *Nat. Electron.* **5**, 452–459 (2022).
19. Ayers, J., Davis, J. L. & Rudolph, A. *Neurotechnology for Biomimetic Robots* (MIT Press, 2002).
20. Webb, B. Robots with insect brains. *Science* **368**, 244–245 (2020).
21. de Ruyter van Steveninck, R. & Laughlin, S. The rate of information transfer at graded-potential synapses. *Nature* **379**, 642–645 (1996).
22. Tuthill, J. C., Nern, A., Holtz, S. L., Rubin, G. M. & Reiser, M. B. Contributions of the 12 neuron classes in the fly lamina to motion vision. *Neuron* **79**, 128–140 (2013).
23. Zheng, L. et al. Network adaptation improves temporal representation of naturalistic stimuli in *Drosophila* eye: I dynamics. *PLoS ONE* **4**, e4307 (2009).
24. Miall, R. The flicker fusion frequencies of six laboratory insects, and the response of the compound eye to mains fluorescent ‘ripple’. *Physiol. Entomol.* **3**, 99–106 (1978).
25. Kelly, D. & Wilson, H. Human flicker sensitivity: two stages of retinal diffusion. *Science* **202**, 896–899 (1978).
26. Uusitalo, R. & Weckström, M. Potentiation in the first visual synapse of the fly compound eye. *J. Neurophysiol.* **83**, 2103–2112 (2000).
27. Nikolaev, A. et al. Network adaptation improves temporal representation of naturalistic stimuli in *Drosophila* eye: II mechanisms. *PLoS ONE* **4**, e4306 (2009).
28. Hu, W., Wang, T., Wang, X. & Han, J. I<sub>h</sub> channels control feedback regulation from amacrine cells to photoreceptors. *PLoS Biol.* **13**, e1002115 (2015).
29. Laughlin, S. B., de Ruyter van Steveninck, R. R. & Anderson, J. C. The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41 (1998).
30. Juusola, M., French, A. S., Uusitalo, R. O. & Weckström, M. Information processing by graded-potential transmission through tonically active synapses. *Trends Neurosci.* **19**, 292–297 (1996).
31. Schuetzenberger, A. & Borst, A. Seeing natural images through the eye of a fly with remote focusing two-photon microscopy. *Iscience* **23**, 101170 (2020).
32. Liu, K. et al. An optoelectronic synapse based on  $\alpha$ -In<sub>2</sub>Se<sub>3</sub> with controllable temporal dynamics for multimode and multiscale reservoir computing. *Nat. Electron.* **5**, 761–773 (2022).
33. Warland, D., Landolfa, M., Miller, J. P. & Bialek, W. in *Analysis and Modeling of Neural Systems* (ed Eeckman, F. H.) 327–333 (Springer, 1992).
34. Jiang, J. et al. Defect engineering for modulating the trap states in 2D photoconductors. *Adv. Mater.* **30**, 1804332 (2018).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2023



## Methods

### Device preparation

Bilayer MoS<sub>2</sub> thin films were grown by the metal–organic chemical vapour deposition method<sup>15</sup>. To prepare the MoS<sub>2</sub> phototransistor, we first fabricated the bottom-gate electrodes (Cr/Au, 3/30 nm) and then prepared the Al<sub>2</sub>O<sub>3</sub> dielectric layer (40 nm) by atomic layer deposition, followed by patterning the source and drain electrodes (Cr/Au, 3/30 nm). The bilayer MoS<sub>2</sub> thin film was transferred onto the substrate and subjected to reactive ion etching to realize a channel with a width of 200  $\mu$ m and a length of 10  $\mu$ m. Finally, a 15-nm-thick Al<sub>2</sub>O<sub>3</sub> encapsulation layer was deposited onto the device to provide electron doping and reduce the variation in the MoS<sub>2</sub> phototransistor arrays after the thermal annealing at moderate temperature ( $\sim$ 200  $^{\circ}$ C).

### Device characterizations

We used a semiconductor analyser system (Keithley 4200-SCS) to conduct the electrical measurements of the MoS<sub>2</sub> devices in an air atmosphere at room temperature. A 660-nm laser (Changchun New Industries) with tunable light intensity was used to generate a constant light that was focused on a specific MoS<sub>2</sub> phototransistor by passing through an optical fibre. Typically, a specific motion was considered to be the combination of four sequential frames. Thus, a temporal sequence of four light pulses was applied to a specific pixel. Each of the phototransistors can sense and process the pixel-level temporal vision sequence independently by means of the local bottom gate (Supplementary Fig. 11). We applied different  $V_g$  to change  $t_0$  to capture different motion scenarios. Accordingly, both light stimulation and readout sampling frequency were set as  $1/t_0$  to obtain distinguishable output data for different stimulation pulses. During the in-sensor motion perception experiments, we tested each device and integrated the output current of each device to map the trajectory of the motion.

### Neuromorphic computing simulations

We demonstrate action recognition of the motion of a solid ball in the real world. We simulated the two artificial vision systems with bioinspired vision sensors and conventional image sensors. Both systems had the same neural network architecture: a  $36 \times 36$  sensor array and a 4-layer perceptron with a  $1,296 \times 100 \times 100 \times 5$  structure. The classic backpropagation algorithm was used for system training and testing. Customized datasets of a solid moving ball were established as training and test datasets. We used a digital camera to capture the video of the ball's movement. The frame rate of the video was 30 frames s<sup>-1</sup>. The resolution of each frame was  $36 \times 36$ . The pixels in the videos had binary values (1 or 0). The spatial information in the video describes the relative positions of pixels, and the temporal information describes the change in a pixel over time. In this way, we captured two datasets for the training and testing of an artificial neural network. The training dataset included five classes of motion (up, down, right, left and blank)

and each class consisted of 600 samples. The testing dataset included the same 5 classes and 50 samples per class. The samples in the testing dataset were different from the samples in the training dataset. Supplementary Note XI provides details of the neuromorphic computing network for action recognition based on the experimentally derived device characteristics.

### Data availability

The datasets are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

### Code availability

The codes used for neuromorphic computing are available from the corresponding authors upon reasonable request.

### Acknowledgements

This work was supported by the Research Grant Council of Hong Kong (CRS\_PolyU502/22), the Shenzhen Science and Technology Innovation Commission (SGDX2020110309540000), the Innovation Technology Fund (ITS/047/20) and The Hong Kong Polytechnic University (1-ZE1T and CD42). J.-H.A. acknowledges support from the National Research Foundation of Korea (NRF-2015R1A3A2A066337).

### Author contributions

Y.C. conceived the concept and supervised the project. J.C. designed the test protocol and performed the experiments. B.J.K. and J.-H.A. fabricated the devices. J.C., Z.W., T.W. and J.Y. analysed the experimental data. Z.Z., Z.W. and Y.Z. performed the simulations. J.C., Z.Z. and Y.C. co-wrote the paper. All the authors discussed the results and commented on the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41565-023-01379-2>.

**Correspondence and requests for materials** should be addressed to Jong-Hyun Ahn or Yang Chai.

**Peer review information** *Nature Nanotechnology* thanks Feng Miao, Guozhen Shen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).