

# CMOS-compatible retinomorphiC Si photodetector for motion detection

Yi WU<sup>1,2</sup>, Wenjie DENG<sup>1,2</sup>, Xiaoqing CHEN<sup>1,2\*</sup>, Jingjie LI<sup>1,2</sup>,  
Songyu LI<sup>1,3</sup> & Yongzhe ZHANG<sup>1,2\*</sup>

<sup>1</sup>Faculty of Materials and Manufacturing, Beijing University of Technology, Beijing 100124, China;

<sup>2</sup>Key Laboratory of Optoelectronics Technology of Education Ministry of China, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

<sup>3</sup>School of Physics, Beihang University, Beijing 100191, China

Received 14 March 2022/Revised 26 June 2022/Accepted 23 September 2022/Published online 4 April 2023

**Abstract** Efficient motion detection is essential for the Internet of Things. However, it suffers from overloading redundant static background information. Inspired by the human visual system, which is efficient in motion detection, we propose a silicon-based retinomorphiC photodetector with a simple metal/insulator/semiconductor (MIS) structure, which is compatible with the complementary metal-oxide-semiconductor (CMOS) industry. In contrast to conventional photodetectors that generate a sustained photocurrent, our retinomorphiC photodetector is sensitive only to the change in light intensity and therefore filters the redundant static background. In addition, it shows logarithmic dependence on the light intensity, which simplifies the contrast ratio measurement. Based on our moving object recognition experiment, after filtering background information, the information to be analyzed is reduced to 27.3%, thereby improving the image recognition efficiency in the subsequent processing tasks. This innovative and industry-compatible retinomorphiC photodetector will facilitate the construction of future efficient motion detection systems.

**Keywords** retina, photodetector, motion detection, bio-inspired, in-sensor computing

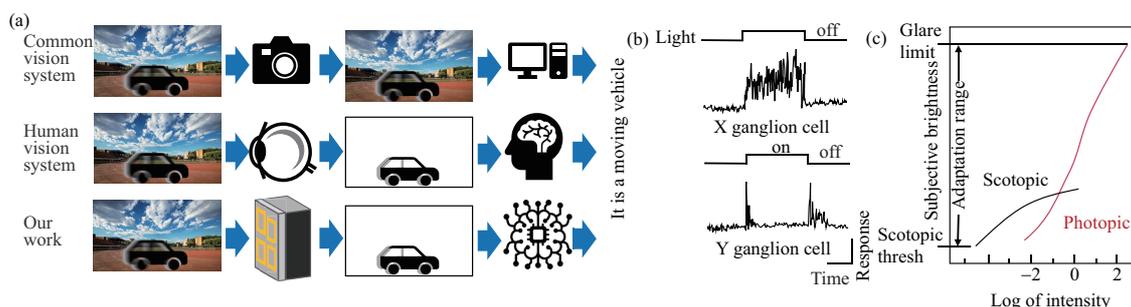
**Citation** Wu Y, Deng W J, Chen X Q, et al. CMOS-compatible retinomorphiC Si photodetector for motion detection. *Sci China Inf Sci*, 2023, 66(6): 162401, <https://doi.org/10.1007/s11432-022-3591-5>

## 1 Introduction

As the artificial intelligence era approaches, the requirement for motion detection is rapidly growing, e.g., in security monitoring, autonomous vehicles, and military reconnaissance [1, 2]. The state-of-the-art hardware system for motion detection is based on a von Neumann architecture and complementary metal-oxide-semiconductor (CMOS) image sensor platform. The sensor unit and processing unit are physically separated and connected via a data bus [3]. In this regard, redundant full image data will inevitably crowd the limited transmission and processing capacity [4]. To improve the efficiency of motion detection, filtering redundant information is necessary. For example, when driving on a highway, relatively static objects could be given lower priority than relatively accelerating or decelerating objects.

Different from the von Neumann architecture CMOS platform, the human visual system is a well-established motion detection system for filtering redundant information [5] when the processing speed of the human brain is slower compared to that of computers [6], as shown in Figure 1(a). For example, a human can easily “sense” a tennis ball at 160 mph speed without really “seeing” it. The biological fundamental for “sensing” and “seeing” is the different types of ganglion cells [7–10], of which the X ganglion cells are responsible for “seeing”, namely, the detection of colorful static images, and the Y ganglion cells are responsible for “sensing”, namely, the detection of motion from the static background, as shown in Figure 1(b). In this way, the information quantity to be transmitted to the brain is significantly reduced, allowing humans to focus on more important moving objects [11, 12].

\* Corresponding author (email: chenxiaoqing@bjut.edu.cn, yzzhang@bjut.edu.cn)



**Figure 1** (Color online) (a) Diagram of the common visual system, the human visual system, and our retinomorph photodetector implementation for motion target recognition. (b) X ganglion cells generate sustained responses to color, while Y ganglion cells generate transient responses to rapid changes in light intensity. (c) The human visual system has a logarithmic dependence on light intensity that makes it adapt to an extraordinarily wide dynamic range of light intensity.

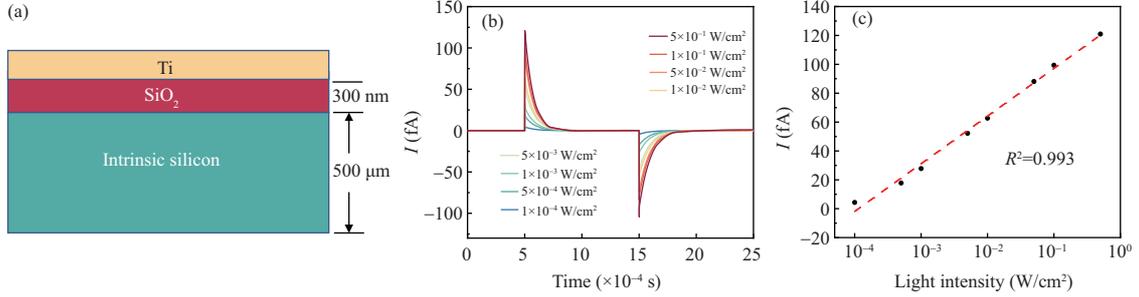
Aside from information filtering, another feature of the human retina is the logarithmic power dependence photoresponse [10]. Compared to the linear photoresponse, the logarithmic photoresponse broadens the dynamic range of light intensity detection, as shown in Figure 1(c), and simplifies the analysis of the ratio before and after the change in light intensity (i.e., contrast ratio), which is more important in the real world.

Various studies have been performed to implement efficient image detection by simulating the human visual system [4, 13–15]. By simulating the structure of the retina, event cameras have advantages to tackling motion detection that is difficult with standard frame-based image sensors, such as high-speed motion monitoring or high-dynamic-range imaging [3, 16, 17]. Up to now, the main strategy to realize the event camera is based on the circuit design consisting of separated photoreceptors, differencing circuits, and comparators, which is inevitably hampered by complicated circuit designs to pursue high performance. Recently, a new architecture—in-sensor computing—was proposed by adopting newly developed semiconductor materials (two-dimensional materials or perovskite) [18–20]. However, newly developed semiconductors often suffer from incompatibility with the present Si industry to achieve rapid commercialization. Therefore, to pursue efficient motion detection, it was necessary to design a new photodetector based on in-sensor computing architecture and compatible with the CMOS.

Here, we fabricated a retinomorph photodetector based on the metal/insulator/semiconductor (MIS) structure and characterized its response to optical stimuli. The device structure is shown in Figure 2(a). Silicon oxide ( $\text{SiO}_2$ ) is sandwiched between intrinsic silicon (i-Si) wafer and titanium electrode, which is equivalent to a chemical capacitor [21–23]. The chemical capacitance arises from the quasi-Fermi level shift upon adding an additional charge. Upon sudden changes in ambient light, the quasi-Fermi level shifts, leading to a change in the capacitance. In this regard, the equilibrium charge density capacity changes, yielding a transient charging/discharging current. By contrast, under static illumination, the MIS structure will exhibit no response due to the presence of its insulating layer. In addition, as the change in the carrier quasi-Fermi energy level of the chemical capacitor shows a logarithmic relationship with the change in the carrier density, we expect that it can simulate the logarithmic optical response of human performance. Aside from these properties, we used silicon as the semiconductor so that it could be perfectly integrated with the current CMOS technology.

## 2 Results and discussion

The photoelectric properties were first evaluated through a simulation using Silvaco. As mentioned above, we expect two features from this silicon MIS device: (1) selectively sensitive to transient light change and (2) logarithmic dependence of photo responsivity on the light intensity. To verify the first feature, a square-wave light was incident from the top to model light changes from the moving object. The simulated current transients under various light intensities are shown in Figure 2(b). As expected, the device generates a current pulse upon a sudden change in the light intensity (positive pulse for light switch-on and negative pulse for light switch-off) while generating zero current when the light intensity is steady (either in dark or under illumination). To verify the second feature, we fit the process with the equation  $y = y_0 + A_0 \exp[-(t - t_0)/\tau]$ , using the parameter  $A_0$  as the current magnitude of the photoresponse, as shown in Figure S2(a). The current amplitude ( $A_0$ ) is plotted against the light intensity, which shows



**Figure 2** (Color online) Simulation of the MIS structure by Silvaco. (a) Device model of the simulation. (b) The temporal response characteristics of the structure demonstrate that the structure has an impulse response to changes in light intensity. (c) The light response shows a logarithmic relationship with the light intensity in the same way as the retina. The dashed line indicates the fitting results, where the fitting residual  $R^2 = 0.993$ .

logarithmic dependence over four orders of magnitude (as shown in Figure 2(c)). More simulation results of the quasi-Fermi energy level, electron-hole concentration, and response current with or without an interface trap are provided in Figures S2(b)–(d). After a successful simulation of the expected device performance, we fabricated devices and device arrays. The optical image of four devices (Ti/SiO<sub>2</sub>/i-Si) is shown in Figure 3(a). The details of the fabrication of the samples are presented in Section 4. The characterization of materials is shown in Figure S4 and Table S1.

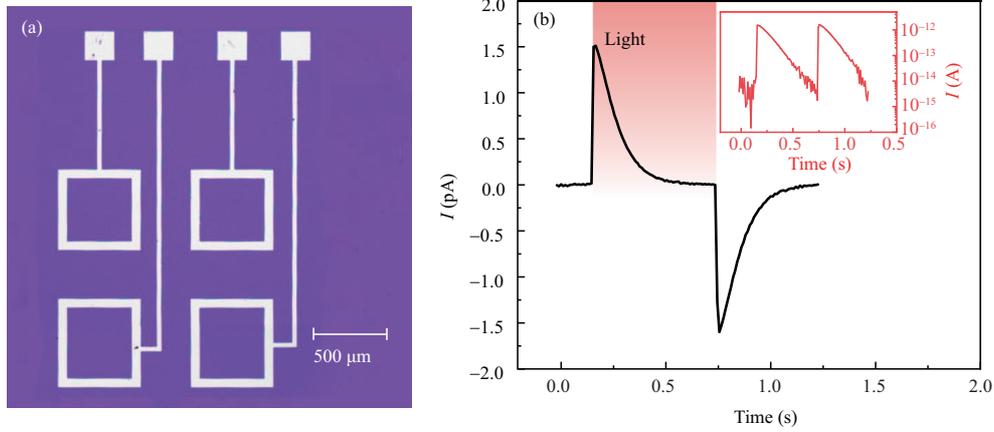
First, we studied the response of a single device to a simple temporal varying light spot to confirm the required device performance. The device current was recorded by grounding the Si electrode. The temporal current response is shown in Figure 3(b). Similar to the simulation results, the current in this system is nearly zero under equilibrium. When a light stimulus (685 nm and a power of 51.2  $\mu$ W) was introduced or removed, the device produced a positive or negative pulse current, respectively. The mechanism of the pulse current generation is described in Note S1. The modulation effect of the voltage was also measured (see Figure S5), which indicates that a larger response photocurrent can be achieved by increasing the voltage. If no special mention, the following tests were measured at 0 V.

The power dependence of the response current transients is shown in Figure 4(a), revealing a logarithmic correlation (Figure 4(b)). Because our semiconductor laser exhibits different wavelength centers at low and high powers, we fitted the experimental data in two segments (see Note S3 for the detailed reason). The logarithmic nature of the response is similar to that of the retina, which implies that the current difference between two points is proportional to the contrast ratio between the two corresponding points in the image, independent of the incident light intensity [24]. As mentioned in the introduction, the advantage of such logarithmic power dependence is that it simplifies the detection of the contrast ratio. To confirm this, we studied the device performance under stepwise increasing light, as shown in Figure 4(c). As expected, a current pulse was generated only when the light intensity changed while the pulse amplitude was proportional to the contrast ratio ( $P_{\text{after}}/P_{\text{before}}$ , Figure 4(d)). Based on this feature, the contrast ratio can be directly obtained without adopting external calculation units.

The mechanism for filtering the static light and detecting change light can be explained by (1). Because the change in the light intensity ( $P$ ) results in the change in the photosensitive chemical capacitance ( $C(V, P)$ ), the equilibrium charge quantity ( $Q$ ) is changed, resulting in a transient current response ( $I$ ). Apparently, the transient current is largely dependent on the rising speed of light intensity, which is the criteria to determine “static light” and “changing light.” To confirm this conclusion, the current responses under the light stimulation of different rising speeds were measured when the square-wave amplitude was fixed at 51.2  $\mu$ W (as shown in Figure 4(e)). The amplitude of the response current is plotted against the rising time in Figure 4(f). Our results show that when the rising time  $< 0.02$  s, the current amplitude remains almost constant, whereas when the rising time exceeds 0.02 s, the current amplitude starts to decrease. When the rising time is 1 s, the current amplitude decreases to  $-3$  dB.

$$I = \frac{dQ}{dt} = V \frac{dC(V, P)}{dt} = V \frac{\partial C(V, P)}{\partial P} \frac{dP}{dt}. \quad (1)$$

Next, after studying the retina response to a temporal varying light spot, we examined the response to a moving light spot, namely, a spatial-temporal varying light spot (Figure 5(a)). For this purpose, two adjacent detectors were employed. Crosstalk between neighboring devices is negligible, as shown in Figure S6. The testing process is shown in Figure S7. The temporal trajectory of the spot movement



**Figure 3** (Color online) Retina-inspired silicon retinomorph detector. (a) Optical image of our designed Ti/SiO<sub>2</sub>/i-Si retina retinomorph detector; (b) photoresponse of a retinomorph detector with a wavelength of 685 nm and power of 51.2 μW; inset: logarithmic plot.

can be obtained according to the temporal relationship between the emergence of the pulse current and the spatial position of the pixel. This phenomenon is similar to the working mechanism of ganglion cells of the human retina.

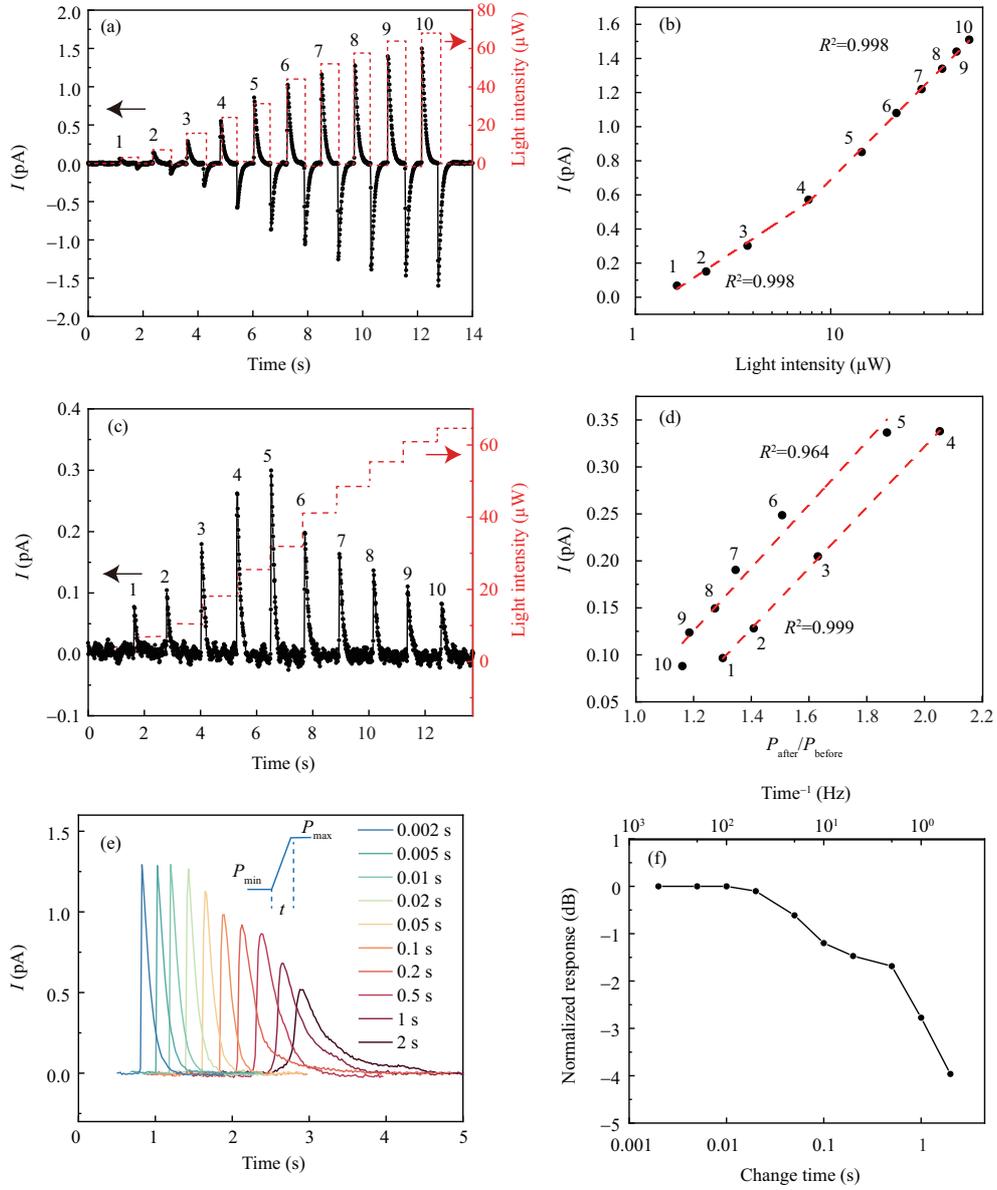
Finally, we studied the retina response to a moving object, which could be modeled as a spatial-temporal varying light image because moving objects generally result in a change in brightness. For this purpose, a 3×3 detector array was employed, and a moving toy vehicle was employed as the moving object. The experimental details can be found in Section 4 and Figure S8. The image of the moving toy vehicle was inputted into the detector array. Because of the difference between the reflected light of the toy vehicle and the background, the light incident to a specific pixel was brightened or darkened when the toy vehicle moved in or out of the view field of the pixel during the toy vehicle's motion. As a result, negative and positive currents could be measured at the front and back of the toy vehicle, respectively, as shown in Figure 5(b). The change in the edge of the toy vehicle before and after the movement is marked by the red dotted line and blue dotted line, respectively. The image output from the detector is similar to the simulation results obtained using the frame difference method [25] in a computer. Details on the frame difference method are presented in Section 4.

As mentioned in Section 1, the advantage of the retinomorph sensor over a regular image sensor is the filtering of redundant information. To quantify the filtering efficiency, information entropy, a concept from information theory, is introduced to analyze the amount of information contained in the raw image and filtered motion images from our retinal detector [26]. The information entropy of the image can be obtained using (2) and the image histogram (i.e., the brightness distribution of the image) [27]. Figure 5(c) shows the image histogram with or without filtering by our detector, where the pixel brightness ranges from 0 to 255.

$$H(m) = -\sum_{i=0}^{L-1} p(m_i) \log_2 p(m_i), \quad \sum_{i=0}^{L-1} p(m_i) = 1, \quad (2)$$

where  $m_i$  and  $L$  indicate the gray value and description, respectively, and  $p(m_i)$  indicates the probability of the occurrence of grayscale  $m_i$  in the image. The results reveal that the information entropy of the images with and without being filtered by our detector is 1.857 and 6.805, respectively. In other words, the amount of information in the image filtered by our detector is only 27.3% of the raw image. Hence, the detector significantly reduces the amount of information in the picture without losing the moving object.

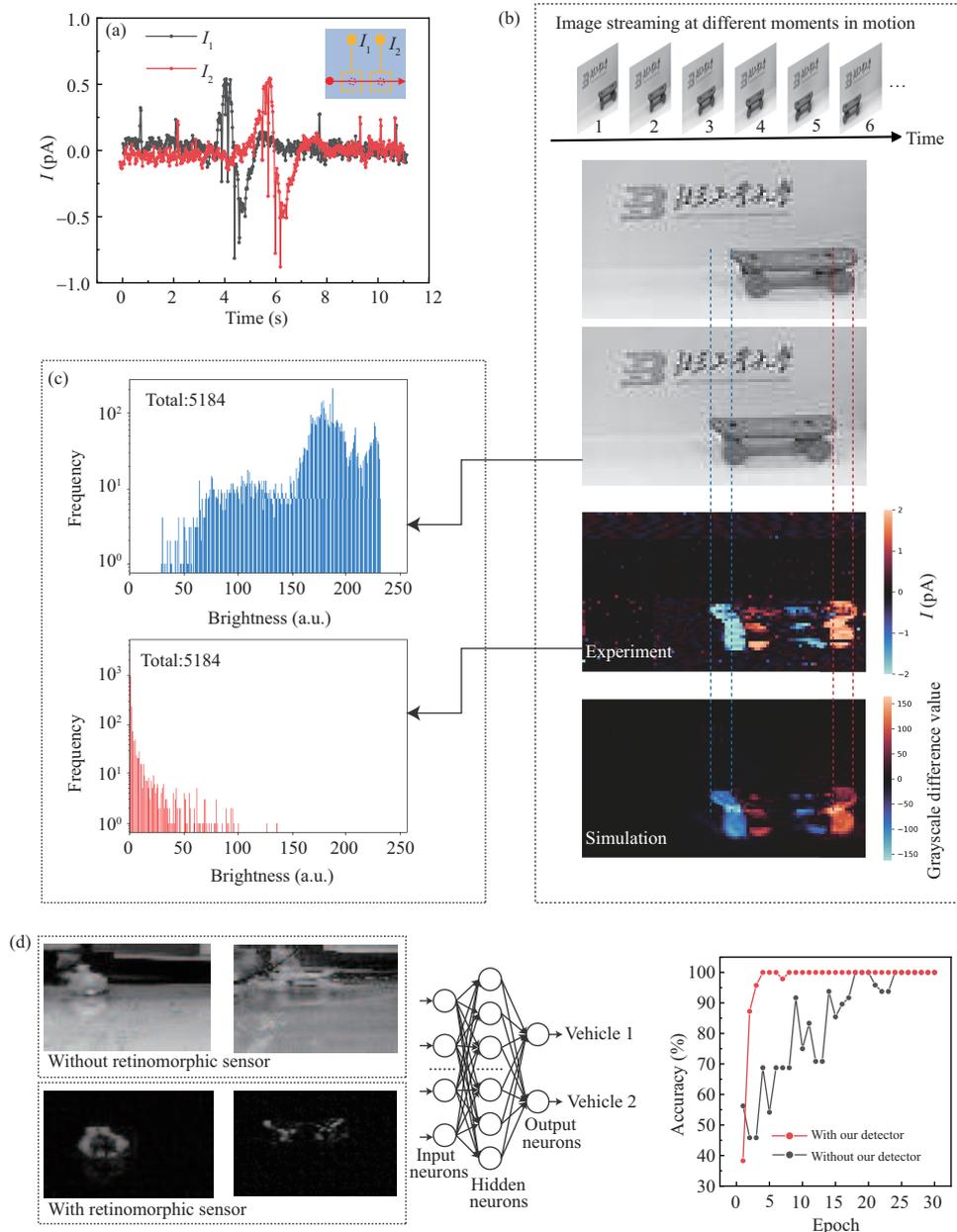
To confirm that redundant information filtering can facilitate further image processing, we tested the accuracy of object recognition from either the raw image or filtered motion image from our retina detector. The mission is to recognize the correct toy vehicle from a database of two toy vehicles. The raw and filtered images of the two toy vehicles are compared in Figure 5(d). With the retinomorph detector, the profile of the car was highlighted, and the complex backgrounds were filtered to black. A three-layer artificial neural network (ANN) was introduced to accomplish the training and recognition of images in the database. We implemented image recognition with the assistance of an ANN with an input layer, a



**Figure 4** (Color online) Photoresponse properties of the silicon retinomorph detector. (a) Current under different power irradiation. (b) Logarithmic variation of the current with the laser power. The dashed lines show the logarithmic fitting curve. (c) Photocurrent under stepwise increasing light. (d) Linear variation of the current with the contrast ratio. The dashed lines show the linear fitting. (e) Current under the light of different variation rates. (f) Relationship between the rising time and photocurrent response. The magnitude of the pulse currents with times less than 0.02 s remains almost constant.

hidden layer, and an output layer. The accuracy gradually improved with increasing training epochs, as shown in Figure 5(d). In particular, to reach 95% recognition accuracy, 20 training epochs were needed when the ANN was trained with the raw image, whereas only four training epochs were required when trained with the filtered image. The results confirm the advantage of our retina detector in analyzing moving objects.

The distinctive optoelectronic response property of our retinomorph photodetector can effectively filter moving objects from the static background, significantly reducing the amount of data to be processed subsequently and providing a new avenue for next-generation machine vision systems [28]. The sensing and processing integrated device structure and CMOS-compatible properties provide the possibility of rapid commercialization in the future.



**Figure 5** (Color online) Illustration of motion detection based on the silicon retinomorph detector. (a) Detection of moving light spots realized with two detectors. (b) Results of the moving toy vehicle under the detection by the  $3 \times 3$  detector arrays. The blue and red dashed lines correspond to the front and rear edges of the moving toy vehicle, respectively. The simulation results are obtained using the frame difference method, and the colors represent the difference in the light intensity (indicated by the gray levels) before and after the motion event. (c) Image histogram with and without filtered by our detector. It represents the frequency of 256 gray levels appearing in the picture. (d) Schematic of the toy vehicle database and ANN recognition process. The database based on the image detected motion toy vehicle results. The ANN includes an input layer, a hidden layer, and an output layer. The recognition accuracy increases with training epochs. Only four epochs are needed to achieve motion target recognition with nearly 100% accuracy.

### 3 Conclusion

In summary, based on the photoresponse properties of the human retina, we propose a silicon-based retinomorph photodetector. The photoresponse of this retinomorph photodetector occurs only when the light intensity changes, which is of a more distinct difference compared to conventional photodetectors that generate sustained photocurrent. The photoresponse of this detector shows a logarithmic relationship with the optical power, making the photocurrent proportional to the contrast ratio and independent of the absolute value of the light intensity. This property has been used to implement a simple detection

of moving spots. Furthermore, we realized the detection of motion toy vehicles because it is silicon-based facilitated integration. The experimental results agree with the computer simulations. The images processed by this retinomorph photodetector maintain the characteristics of original moving objects and significantly reduce the amount of information contained in the images, resulting in a significant reduction in training for subsequent recognition using ANNs. Our work provides new insights into the application of intelligent photodetectors to the smart Internet of Things and artificial retinas.

## 4 Materials and method

**Simulations of the devices.** Simulations were performed using the ATLAS module of the Silvaco TCAD tool. The two-dimensional structure cross section was defined with a 300 nm layer of silicon dioxide and a 500  $\mu\text{m}$  layer of silicon. The work function of the top gate was set to that of Ti (4.33 eV). The light source conditions were consistent with those in the experiment.

**Fabrication of the silicon retinomorph detector.** The intrinsic silicon substrates were purchased from Nanjing MKNANO Tech. Co., Ltd. ([www.mukenano.com](http://www.mukenano.com)). Before further deposition, the wafers were cleaned with acetone, anhydrous ethanol, and deionized water in sequence. Then, 10 nm of Ti and 70 nm of Au were deposited as electrodes through the ultraviolet lithography technology (SUSS MJB4) and e-beam evaporation (HHV FL400). Finally, the lift-off technique was used to complete the final device.

**Characterization of the devices.** The optical images of the fabricated devices were characterized via optical microscopy (Olympus BX 51). The semiconductor parameter analyzer (Agilent B1500A) was connected to the probe station for electrical characterization in ambient air. In the photoresponse measurements, an LED laser modulated with a waveform generator and a supercontinuum light source (SuperK Extreme, NKT Photonics) was used to illuminate the device. The Raman spectra were measured by a confocal Raman microscope (Witec Alpha 300) with 532 nm laser excitation to characterize the materials. Scanning electron microscopy (FEI Inspect F50), energy-dispersive spectroscopy (EDAX super octane), and X-ray photoelectron spectroscopy (Thermo Fisher, ESCALAB 250Xi) were used to characterize the materials.

**Frame difference method.** This method detects moving objects by comparing the difference between the existing frame and the reference frame. Each frame corresponds to a matrix of grayscale values, and because object motion causes a change in light intensity (grayscale values), the two matrixes were subtracted to detect which pixels contain moving objects. In the simulation process, we first took a video (30 fps) of the motion of a toy vehicle using a camera. Then, we used Python to extract two frames of the video and subtract them, and the final image is a motion image detected by the frame difference method.

**Detection of the moving toy vehicle.** We simulated the pixel values in the toy vehicle motion image by the photoresponses of the device under different laser irradiation and thus realized the mimicking of the real motion image detection. We downsampled the moving toy vehicle image captured by a digital camera into  $96 \times 54$ , and the pixel value was compressed to a range from 0 to 255. A waveform generator feeds the pixel value into the laser in chronological order. As illustrated in Figure S8, pixel values at neighboring moments were arranged together, and calibration signals were inserted for good signal separation. Then, we measured the current from the retinomorph detector and reorganized the current data into images.

**Recognition of the moving toy vehicles.** A three-layer ANN was built by MATLAB to implement the recognition of the toy vehicle motion. The backpropagation algorithm was used for the training and recognition processes. The activation function in the hidden layer and output layer is Softmax. The learning rate was set at a fixed value of 0.01. The images filtered by the detector were directly connected to the input layer in the neural network.

**Acknowledgements** This work was supported by Beijing Natural Science Foundation (Grant Nos. JQ20027, 4222065), National Natural Science Foundation of China (Grant No. 61922005), NSAF Joint Fund (Grant No. U1930105), and National Key Research and Development Program of China (Grant No. 2021YFA0715602).

**Supporting information** Appendixes A–C, Notes S1–S3, Table S1, and Figures S1–S8. The supporting information is available online at [info.scichina.com](http://info.scichina.com) and [link.springer.com](http://link.springer.com). The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

## References

- 1 Cutler R, Davis L S. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans Pattern Anal Machine Intell*, 2000, 22: 781–796

- 2 Chen T, Lu S. Object-level motion detection from moving cameras. *IEEE Trans Circuits Syst Video Technol*, 2017, 27: 2333–2343
- 3 Gallego G, Delbruck T, Orchard G, et al. Event-based vision: a survey. *IEEE Trans Pattern Anal Mach Intell*, 2022, 44: 154–180
- 4 Zhou F, Zhou Z, Chen J, et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat Nanotechnol*, 2019, 14: 776–782
- 5 Ölveczky B P, Baccus S A, Meister M. Segregation of object and background motion in the retina. *Nature*, 2003, 423: 401–408
- 6 Zhou F, Chai Y. Near-sensor and in-sensor computing. *Nat Electron*, 2020, 3: 664–671
- 7 Erskine L, Herrera E. Connecting the retina to the brain. *ASN Neuro*, 2014, 6: 175909141456210
- 8 Masland R H. The fundamental plan of the retina. *Nat Neurosci*, 2001, 4: 877–886
- 9 Shapley R M, Victor J D. Nonlinear spatial summation and the contrast gain control of cat retinal ganglion cells. *J Physiol*, 1979, 290: 141–161
- 10 Gonzalez R C, Woods R E, Masters B R. Digital image processing, third edition. *J Biomed Opt*, 2009, 14: 029901
- 11 Wang G Y, Liets L C, Chalupa L M. Unique functional properties of on and off pathways in the developing mammalian retina. *J Neurosci*, 2001, 21: 4310–4317
- 12 Enroth-Cugell C, Robson J G. The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol*, 1966, 187: 517–552
- 13 Lange E, Funatsu E, Hara K, et al. Artificial retina devices-fast front ends for neural image processing systems. In: *Proceedings of International Joint Conference on Neural Networks*, 1993
- 14 Wang C Y, Liang S J, Wang S, et al. Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor. *Sci Adv*, 2020, 6: eaba6173
- 15 Liao F, Zhou Z, Kim B J, et al. Bioinspired in-sensor visual adaptation for accurate perception. *Nat Electron*, 2022, 5: 84–91
- 16 Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15  $\mu$ s latency asynchronous temporal contrast vision sensor. *IEEE J Solid-State Circuits*, 2008, 43: 566–576
- 17 Posch C, Serrano-Gotarredona T, Linares-Barranco B, et al. Retinomorph event-based vision sensors: bioinspired cameras with spiking output. *Proc IEEE*, 2014, 102: 1470–1484
- 18 Zhang Z, Wang S, Liu C, et al. All-in-one two-dimensional retinomorph hardware device for motion detection and recognition. *Nat Nanotechnol*, 2021, 17: 27–32
- 19 Herrera C T, Labram J G. A perovskite retinomorph sensor. *Appl Phys Lett*, 2020, 117: 233501
- 20 Herrera C T, Labram J G. An organic retinomorph sensor. *ACS Appl Electron Mater*, 2022, 4: 92–98
- 21 Luryi S. Quantum capacitance devices. *Appl Phys Lett*, 1988, 52: 501–503
- 22 Bisquert J. Chemical capacitance of nanostructured semiconductors: its origin and significance for nanocomposite solar cells. *Phys Chem Chem Phys*, 2003, 5: 5360
- 23 Almora O, Garcia-Belmonte G. Light capacitances in silicon and perovskite solar cells. *Sol Energy*, 2019, 189: 103–110
- 24 Mead C A, Mahowald M A. A silicon model of early visual processing. *Neural Networks*, 1988, 1: 91–97
- 25 Neri A, Colonnese S, Russo G, et al. Automatic moving object and background separation. *Signal Processing*, 1998, 66: 219–232
- 26 Shannon C E. A mathematical theory of communication. *Bell Syst Technical J*, 1948, 27: 379–423
- 27 Dong J, Wu G, Yang T, et al. The improved image scrambling algorithm for the wireless image transmission systems of UAVs. *Sensors*, 2018, 18: 3430
- 28 Wu P, He T, Zhu H, et al. Next-generation machine vision systems incorporating two-dimensional materials: progress and perspectives. *InfoMat*, 2022, 4: e12275