

<https://doi.org/10.1038/s44335-025-00031-7>

# Bio-inspired optoelectronic devices and systems for energy-efficient in-sensor computing

Check for updates

Xiaoting Wang<sup>1,2,5</sup>, Heyi Huang<sup>1,3,5</sup>, Jianshi Tang<sup>1</sup> ✉, Ruofei Hu<sup>1</sup>, Yiwei Du<sup>1</sup>, Yuyan Wang<sup>1</sup>, Bin Gao<sup>1,4</sup>, He Qian<sup>1</sup> & Huaqiang Wu<sup>1,4</sup> ✉

AI-driven machine vision faces critical energy and latency challenges from exponential growth in visual data. This review proposes bio-inspired energy-efficient in-sensor computing utilizing emerging optoelectronic memristors, examining neural network architectures (fully connected/convolutional, recurrent, and spiking neural networks) for static, motion, and event-driven processing that directly emulates biological visual pathways. Critical integration challenges and strategic roadmaps are systematically analysed to achieve cortex-level energy efficiency in next-generation vision systems.

AI-driven machine vision has rapidly developed, enabling artificial systems to “see” and extract valuable information from images and videos<sup>1–5</sup>. In sectors like autonomous vehicles, these systems perform complex vision tasks<sup>6</sup>, such as driving on freeways, avoiding collisions, following lanes, and automatically overtaking slower vehicles. Current implementations, however, rely heavily on the deployment of numerous sensors, with image sensors (using conventional CMOS imagers or CCD cameras) for capturing visual data, infrared sensors for perimeter detection, and other sensors like radar or ultrasonic sensors for precise distance detection. Projections indicate that by 2032, the data generated from such sensors may escalate to approximately  $10^{27}$  bytes per year, with visual systems contributing 80% of this volume<sup>7</sup>. And this surge of unprocessed data often contains redundancies, due to frame-based imaging modes, which poses critical challenges for data-intensive and delay-sensitive applications (e.g., robotic navigation)<sup>8</sup>. Furthermore, in current von Neumann architecture-based systems (e.g., software neural networks or CMOS synaptic circuits), the analog-to-digital conversion of massive visual data streams requires separate memory units for storage and processing. This physical separation of sensing, computing, and memory units results in energy inefficiency and communication latency<sup>9</sup>.

The development of energy-efficient and intelligent alternative technologies has become critical for overcoming the “von Neumann bottleneck”. Biological visual systems demonstrate superior efficiency (femtojoules per operation, fJ/operation) through four mechanisms: (1) hierarchical biostructure, which compresses data through edge detection and dynamic range adjustment; (2) task-specific parallel pathways for motion, color, and intensity processing; (3) event-driven encoding, and (4)

retinal preprocessing that filters 80% of redundant data prior to neural transmission<sup>10,11</sup>. These principles have inspired “in-sensor computing”<sup>12–15</sup>, which shifts partial computational from centralized processors to sensory terminals<sup>16</sup>. Optoelectronic memristors (OEMs) exemplify this paradigm<sup>17</sup>, integrating optical sensing and computation to directly compress multi-dimensional data (intensity, wavelength, polarization) into machine-interpretable patterns<sup>18–21</sup>. Volatile OEMs further enable transient storage for spatiotemporal processing. To extract higher-dimensional optical features (spectral, polarization, phase), advanced strategies employ metasurfaces or specialized optoelectronic devices. In the spectral dimension, materials with varying transmission coefficients act as wavelength-selective bandpass filters to extract color features. For the polarization dimension, highly anisotropic materials or chiral plasmonic metasurfaces enhance linear/circular polarization detection, even reconstructing full Stokes parameters in extreme environments (e.g., underwater)<sup>22–24</sup>. In the phase dimension, metasurfaces with pre-engineered phase profiles act as modulators to extract the specific phase information<sup>25</sup>.

In this review, we will begin by exploring the pathways of the human visual system and their relevance to bio-inspired neural networks designed in in-sensor computing. The basic device structures of OEMs, including two-terminal and three-terminal configurations, will also be discussed, along with their operational mechanisms. Building on these memristors, we will provide a detailed summary of recent developments in bio-inspired neural architectures, focusing on three primary areas: (1) in-sensor fully connected networks (FCNs)/convolutional neural networks (CNNs) for static image recognition, (2) in-sensor recurrent neural networks (RNNs) for motion recognition, and (3) in-sensor spiking neural networks (SNNs)

<sup>1</sup>School of Integrated Circuits, Beijing Innovation Center for Integrated Circuits, Tsinghua University, Beijing, China. <sup>2</sup>Key Laboratory of Optoelectronics Technology of Education, School of Information Science and Technology, Beijing University of Technology, Beijing, China. <sup>3</sup>Integrated Circuit Advanced Process R&D Center and Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China.

<sup>4</sup>International Innovation Center of Tsinghua University, Shanghai, China. <sup>5</sup>These authors contributed equally: Xiaoting Wang, Heyi Huang.

✉ e-mail: [jtang@tsinghua.edu.cn](mailto:jtang@tsinghua.edu.cn); [wuhq@tsinghua.edu.cn](mailto:wuhq@tsinghua.edu.cn)

for event-based dynamic information. Furthermore, we will consider potential deployment strategies for these systems using advanced manufacturing technologies, and outline a strategic roadmap that highlights the key challenges facing the development of in-sensor computing.

**Biological inspiration for in-sensor computing**

The intricate interplay between the retina and the visual cortex forms the basis of the brain’s hierarchical, compact, and adaptable learning processes. Studies show that visual information is transmitted and processed through complex hierarchical pathways in the brain<sup>26–29</sup>. Visual signals are initially transduced by photoreceptors in the retina, where rods specialize in low-light vision and cones facilitate color and high-resolution visual tasks. These signals are conveyed via the optic nerve to the lateral geniculate nucleus (LGN), which acts as a crucial relay and integrative hub for visual information before it reaches the primary visual cortex (V1). In V1, spatial and temporal features such as edge orientation, contrast, and motion are further analyzed. From V1, the visual stream diverges into specialized regions, including V2 (secondary visual cortex), V3 (tertiary visual cortex, critical for form-motion integration), MT (middle temporal area), V4, and IT (inferotemporal cortex)<sup>30</sup>.

Different neural pathways within this system specialize in different aspects of visual processing (Fig. 1). The ventral stream (“what pathway”), involving V1, V2, V4, and IT, excels in object recognition by analyzing shape, color, and texture. In contrast, the dorsal stream (“how pathway”) processes motion and spatial relationships through V1, V2, MT, and the parietal cortex, enabling real-time navigation and action planning<sup>31</sup>. This dual-pathway mechanism serves as a biological blueprint for multi-task decision-making. For example, in robotic vision, ventral-like algorithms identify object features (e.g., shape), while dorsal-like models analyze the spatial position and posture, improving operational precision. Similarly, autonomous vehicles integrate ventral-like static recognition (e.g., traffic lights) with dorsal-like dynamic prediction (e.g., pedestrian trajectories) using hierarchical algorithms that fuse low-level edge detection and high-level semantic mapping. Inspired by these biological systems, researchers are developing hardware that mimics these neural hierarchies, spanning low-level sensory processing (e.g., object recognition, motion detection) to high-level cognitive integration (e.g., decision-making, trajectory prediction). This hierarchical design enables simultaneous static and dynamic visual processing across temporal sequences and multiple targets, achieving biological-level efficiency in complex environments<sup>30,32,33</sup>.

**Optoelectronic device architectures for in-sensor computing**

The emulation of biological vision systems in hardware requires physical models at two different levels. In this section, we will first discuss individual components such as artificial synapses and neurons.

**Photoreceptors**

The human retina is a complex biological image detector that integrates various specialized cell types, including photoreceptors (rods and cones),

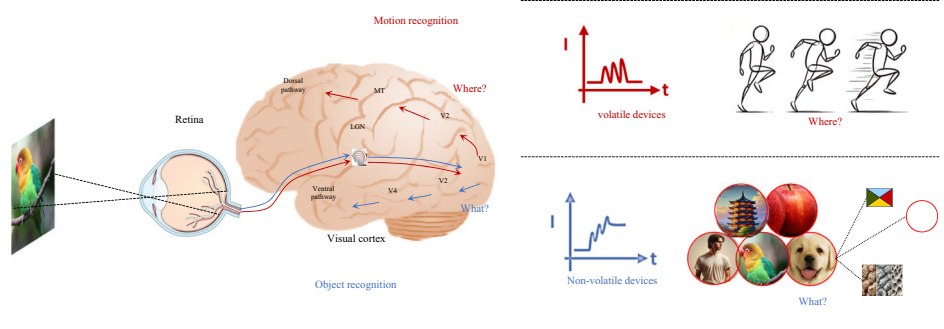
bipolar cells, amacrine cells, and ganglion cells. Photoreceptors serve as the primary phototransducers, converting light signals into neuro-electric pulses through biochemical reactions<sup>34</sup>. These light-sensitive cells form two functional systems: the cone-mediated photopic system for daylight vision and the rod-dominated scotopic system for low-light detection. The photopic system features approximately 6 million cone cells densely packed in the fovea, enabling color discrimination and high-resolution vision under bright illumination. Three spectral subtypes S/M/L cones collectively allow for the recognition of over 1 million color variations through trichromatic encoding. In contrast, the scotopic system employs 90 to 120 million rod cells distributed peripherally, which are optimized for detecting faint light and motion but lacking color sensitivity. Retinal signal processing involves a hierarchical refinement of visual information. In this process, bipolar cells integrate information from multiple photoreceptors and relay the outputs to ganglion cells, while amacrine cells modulate spatial and temporal features through lateral connections. This neural network extracts essential visual parameters, such as luminance, contrast, edge orientation, and motion direction. Once these parameters are processed, ganglion cells encode the information into action potentials. These compressed neural signals are then transmitted via the optic nerve to the visual cortex for higher-order interpretation.

To ensure photoreceptors to adapt to varying lighting conditions, the retina utilizes two types of biological visual adaptation: photopic adaptation and scotopic adaptation. When transitioning from bright to dark environments, the retina initially struggles to detect objects because cone cells require strong light to function. However, through the scotopic adaptation process, the rod cells gradually become activated and increase light sensitivity to low light, allowing for the recognition of dark objects. Conversely, in bright light conditions, the retina may be temporarily overwhelmed by bright objects, making it difficult to discern details. Through photopic adaptation, the visual system can gradually adjust by suppressing ambient light noise and enhancing the sensitivity of cone cells, enabling the recognition of bright objects over time. This dual adaptation optimizes dynamic range, allowing vision from starlight to daylight<sup>35,36</sup>. Recently, Chai et al. developed MoS<sub>2</sub> phototransistors with an environment-adaptive image-sensing capability. Many charge trap sites in the channel/insulator interface region were intentionally introduced to enable the photopic and scotopic sensing operations. The visual adaptation mode could be activated by applying a constant gate bias to the gate terminal<sup>12</sup>. However, the major limitation of currently developed adaptive neuromorphic devices is that they cannot automatically switch their adaptive mode according to the ambient lighting environment. In addition to photoreceptor-like devices for adaptive light detection, integrating these photoreceptors with ganglion cells can enhance functionalities like contrast enhancement and various preprocessing tasks.

**Optoelectronic neuron**

The human retina operates as a distributed computing architecture, where heterogeneous neurons encode visual information through spatiotemporal spike patterns. Mimicking this paradigm, optoelectronic Leaky Integrate-

**Fig. 1 | Parallel processing in visual pathways.** The dorsal pathway (in red) specializes in motion recognition and utilizes volatile devices to simulate and track object location and movement. In contrast, the ventral pathway (in blue) focuses on object recognition, employing non-volatile devices to identify and classify static features.



and-Fire (LIF) neurons implement biomimetic spike-based processing, enabling energy-efficient neuromorphic vision systems<sup>37,38</sup>. The fundamental computation of LIF neurons involves the conversion of input signals into spike patterns, which is widely applied in SNNs. Spike-based processing allows for sparse and efficient information transmission and compression. The conventional LIF model incorporates four key components: threshold-driven firing, “all-or-nothing” property, refractory period, and frequency-dependent tunability. When the membrane potential exceeds a certain threshold, these neurons generate a single electrical signal called a “spike”, with the intensity and duration of the spike reflecting the temporal attributes of the input signal<sup>39</sup>. In optoelectronic LIF neurons, optical signals can be converted into electronic spikes, and the firing frequency of LIF neurons can be modulated in different optical dimensions, thereby promoting the nervous ability to encode visual information<sup>40</sup>.

The LIF neuron model can be implemented using a hybrid circuit composed of threshold-switching memristors (TSM), capacitors, and comparators<sup>40,41</sup>. This model is composed of simple units capable of replicating the spiking behavior of the neurons. A critical component of artificial optoelectronic LIF neurons is a threshold switching memory. The input current charges a parallel capacitance and determines whether a spike is generated by reaching the membrane potential threshold<sup>42</sup>. The TSM switches from high-resistance state (HRS) to low-resistance state (LRS) when the voltage exceeds the threshold ( $V_{th}$ )<sup>43</sup>, and reverts to HRS when the voltage drops below the holding voltage ( $V_{hold}$ ). The transition generates a discharge spike. The integration of threshold switch (ITO/Ta<sub>2</sub>O<sub>5</sub>/Ag) and photosensitive indium–gallium–zinc-oxide (IGZO)-based photoresistor simulates output spike characteristics under light modulation. In a study by Wan et al., LIF neurons responding to optical signals were implemented using ITO/Ta<sub>2</sub>O<sub>5</sub>/Ag/IGZO/ITO device<sup>42</sup>. This device demonstrated the translation of external optical stimuli into spike potentials and achieved low energy consumption of ~400 picowatts per spike. These results reveal the great potential of such devices for constructing an artificial visual system with high energy efficiency and high biological plausibility.

It is important to emphasize that the LIF model replicates only the leaky integrate and fire characteristic of a neuron. Implementing the function of a neuron with a single ion channel requires a similar unit, but to mimic the various ion channels present in the membrane, multiple units must be coupled<sup>44</sup>. The H–H neuron system can emulate a more complex neuronal behavior. From a circuit theoretic standpoint, the Na and K ion channels of the Hodgkin–Huxley model are mathematically equivalent to two distinct memristors<sup>45</sup>.

### Optoelectronic Synapses

Neurons transmit signals through synapses, which can either enhance or inhibit the signal depending on synaptic weights<sup>46</sup>. In optoelectronic neuromorphic computing, devices adjust their photoresponsivity and conductance in response to light and electric stimuli, facilitating updates to synaptic weights<sup>47,48</sup>. Current implementations include two-terminal and three-terminal optoelectronic devices, differing in their weight tuning and read paths.

**Two-terminal optoelectronic synapses.** Two-terminal devices, particularly the metal–semiconductor–metal (MSM) vertical sandwich structure, are widely used in electrical neuromorphic systems. Two-terminal vertical OEMs utilize the same terminals for weight adjustment and reading, making them suitable for high-density arrays due to their simplicity and compactness. We have categorized five main types of two-terminal devices based on their structures and mechanisms, analyzing their main behavior, pros, and cons in application:

**Ferroelectric Control Mechanism:** In this mechanism, the photoresponsivity of devices can be manipulated through ferroelectric polarization (Fig. 2a). The polarization alters the barrier height and the built-in electric field, further affecting the device’s photoresponsivity. This control ability allows for the regulation of both positive and negative photoresponsivity. Thus, the ferroelectric device array can perform an in-situ

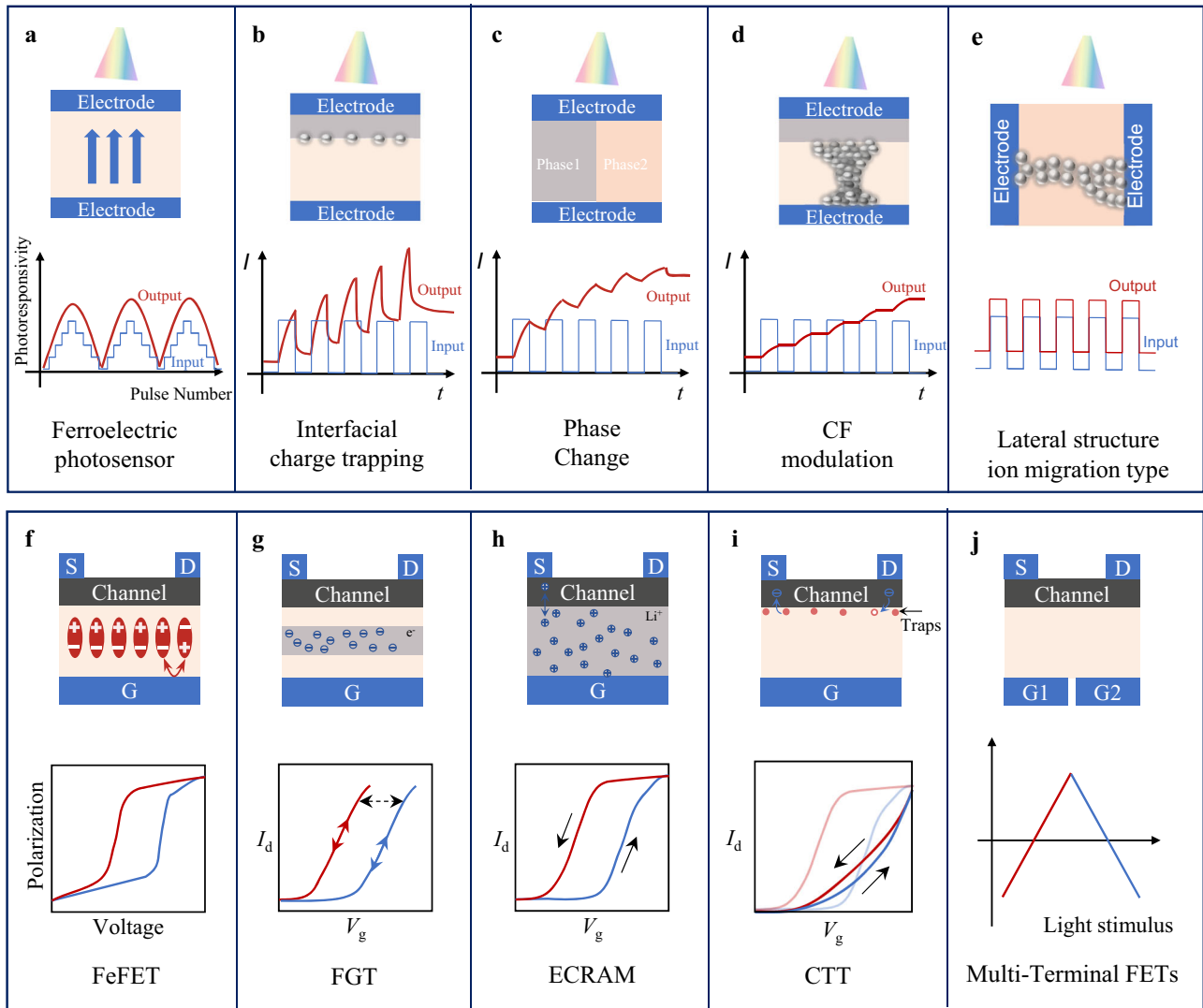
matrix-vector multiplication operation between an input image and a photoresponsivity matrix, enabling object recognition in FCNs. The primary advantage lies in the non-volatility of switched ferroelectric domains combined with rapid switching speeds and low energy consumption, making them suitable for in-sensor computing. For example, a Pt/Pb(Zr<sub>0.2</sub>Ti<sub>0.8</sub>)O<sub>3</sub> (PZT)/SrRuO<sub>3</sub> (SRO) heterostructure can achieve a linear multilevel photoresponse intensity by altering the polarization direction, making it suitable for CNNs<sup>49</sup>. However, the presence of numerous defects within and at the interfaces of ferroelectric materials challenges reliability, and most ferroelectric materials are difficult to integrate with CMOS technology for large-scale applications, except for hafnium-based ferroelectrics.

**Interface Charge Capture Mechanism:** This mechanism involves the capture of photo-generated carriers by the interface barrier and defect vacancies (Fig. 2b). When light is incident on the device, photo-generated carriers are trapped at the interface by the barrier and defects, leading to conductivity modulation. For example, oxide heterojunction structures such as ITO/CeO<sub>2-x</sub>/AlO<sub>y</sub>/Al<sup>50</sup>, ITO/ZnO<sub>1-x</sub>/AlO<sub>y</sub>/Al<sup>51</sup> can modulate their conductance by altering the interface trap and de-trapping states through light modulation. However, due to the localized effect of the density of states, conductivity modulation tends to saturate. To overcome these limitations, researchers are exploring new materials and device structures to enhance the stability and response speed of the devices. Devices based on interface traps are easy to fabricate and have a wide response wavelength range and fast response rates. However, due to the difficult-to-control distribution of defect states, these devices suffer from poor uniformity and relaxation phenomena, making them unsuitable for nonvolatile storage but ideal for dynamic temporal tasks.

**Photo-Induced Phase Change Mechanism (PCM):** In this mechanism, the conductivity of the device can be controlled via photo-induced amorphous/crystalline phase changes (Fig. 2c). By utilizing light of different wavelengths, frequencies, and intensities, precise control over device conductivity can be achieved. For example, MoO<sub>3</sub> can undergo phase changes under light stimulus, allowing for the modulation of its conductance<sup>17</sup>. However, this phase transition process typically requires high light energy consumption in practical applications. To reduce energy consumption, researchers are seeking more efficient photo-induced phase change materials and developing new device structures to lower energy consumptions. Despite demonstrating excellent multi-level storage capabilities, practical implementation faces energy efficiency challenges due to high photon energy requirements, particularly in visible spectrum applications.

**Conductive Filamentary (CF) Optoelectronic Devices:** These devices rely on oxygen vacancies in the conductive filament (Fig. 2d). When light stimulates the device, oxygen vacancies in the filament are ionized, generating more positive vacancies. These vacancies can assist in hopping conduction of electrons, thereby promoting control over the device’s conductivity. This mechanism provides a unique way to control electron transport via light stimulation, offering fresh insights for the development of novel optoelectronic devices. For example, ZnO-based conductive filament states can be modulated by light due to their resistive switching and photosensitive properties<sup>52,53</sup>. Researchers are investigating ways to optimize the structure and materials of the conductive filament to enhance the performance and stability of the devices. The advantage of CF devices is their compatibility with large-scale CMOS integration, wide wavelength response range, and fast response rates. However, the randomness in filament formation challenges the uniformity and reliability of the devices.

In addition, there are planar two-terminal optoelectronic devices that utilize an ion migration mechanism. Under optoelectronic modulation, ionic migration can dynamically form conductive channels, altering carrier distribution and interface barriers. This enables the device to exhibit programmable optoelectronic response characteristics, effectively emulating the plasticity of neural synapses (Fig. 2e). A typical example is the planar two-terminal MSM device constructed using two-dimensional (2D) metal sulfides. In this device, sulfur vacancy migration behavior is precisely controlled by an electric field, actively modulating the metal–semiconductor



**Fig. 2 | Bio-inspired optoelectronic devices.** Two-terminal devices: Ferroelectric-optoelectronic devices (a) whose photoresponsivity can be modulated through ferroelectric pulse polarization. Interface trapping devices (b) capable of enhancing response current under light exposure<sup>51,111</sup>. Photo-induced phase-change devices (c) where conductivity can non-volatily increase with light exposure<sup>112</sup>. CF optoelectronic devices (d) in which photo-generated carriers can control the charge of oxygen vacancies, thereby non-volatily modulating their conductivity<sup>53,113</sup>. Planar two-terminal optoelectronic devices (e) utilizing ion migration mechanisms.

terminal devices: FeFETs (f) utilize partial polarization switching within the ferroelectric gate oxide to change conductance. FGTs (g) operate by trapping charge on a floating gate to control the threshold voltage. Conductance tuning of ECRAM (h) is based on the motion of  $\text{Li}^+$  ions between the electrolyte and channel. CTTs (i) enable precise modulation of electrical characteristics by trapping and releasing photo-generated carriers. Multi-Terminal FETs with a double-gate design (j) for reconfigurable photoresponsivity.

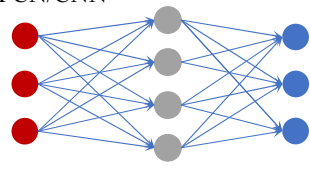
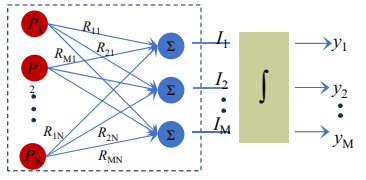
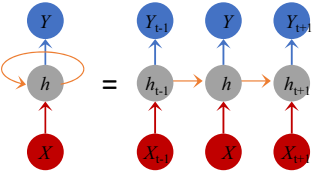
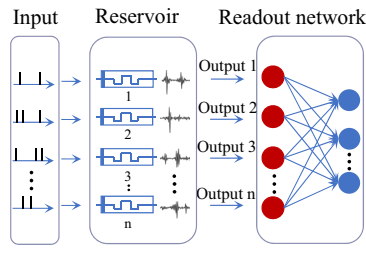
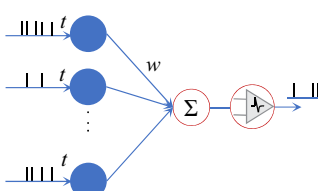
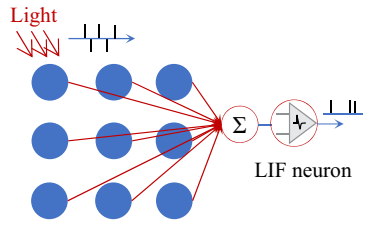
contact barrier. This ion migration-induced dynamic barrier adjustment mechanism drives non-volatile changes the photocurrent, successfully mimicking the synaptic weight update functionality in CNNs. The advantages of planar migration devices include rich modulation behavior, but they suffer from uneven defect state distribution, reliability issues, and larger device sizes, making them difficult to fabricate with CMOS BEOL<sup>54,55</sup>.

**Three-terminal/multi-terminal optoelectronic synapses.** Three-terminal field-effect transistors (FETs) introduce an additional gate terminal, which allows for independent control of the weight tuning process without interfering with the reading path<sup>56</sup>. Thus, three-terminal/multi-terminal OEMs simultaneously perform signal transmission and self-learning, facilitating their use in advanced neuromorphic engineering that mimics the brain, which is challenging for two-terminal devices. Their operation is primarily based on four mechanisms:

Ferroelectric FETs (FeFETs) and floating-gate transistors (FGTs) are advanced technologies for non-volatile photoresponse control. FeFETs

integrate a ferroelectric gate dielectric (e.g., poly(vinylidene fluoride) (PVDF),  $\text{Pb}(\text{Zr},\text{Ti})\text{O}_3$  (PZT),  $\text{LiNbO}_3$ , or zirconium-doped hafnia (HZO), whose intrinsic dipole polarization can be reversibly reoriented via gate pulses<sup>57,58</sup>. When subjected to a gate electric field, partial alignment of ferroelectric domains induces bound polarization charges at the dielectric-channel interface. These charges modulate the channel conductivity through electrostatic screening (Fig. 2f), inducing a non-volatile conductance state.

FGTs achieve non-volatile optical conductance updates through an embedded FG layer within the dielectric layer (Fig. 2g)<sup>59,60</sup>. This architecture vertically integrates a thick blocking dielectric, photoresponsive semiconducting FGs, an ultrathin tunneling dielectric (e.g.,  $\text{Al}_2\text{O}_3$ ,  $\text{HfO}_2$ ), and a semiconducting channel with a wider bandgap than the FG material. When sub-bandgap light is absorbed by FG materials (e.g., perovskite quantum dots (QDs), black phosphorus (BP) QDs, optimized for visible-to-near-infrared sensitivity)<sup>61</sup>, photogenerated electron-hole pairs separate due to the aligned band structure: electrons inject into the channel to enhance

Architecture	Building blocks	Key features	Implementation example
<b>a</b> FCN/CNN 	<ul style="list-style-type: none"> <li><span style="color: blue;">●</span> Input neurons</li> <li><span style="color: blue;">→</span> Weights</li> <li><span style="color: grey;">●</span> Hidden neurons</li> <li><span style="color: red;">●</span> Output neurons</li> </ul>	<ul style="list-style-type: none"> <li>• All neurons connect</li> <li>• Classification and regression tasks</li> </ul> Pros: <ul style="list-style-type: none"> <li>• Simple</li> <li>• Good for structured data</li> </ul> Cons: <ul style="list-style-type: none"> <li>• Inefficient for sequential data</li> <li>• High computation cost</li> </ul>	
<b>b</b> RNN 	<ul style="list-style-type: none"> <li><span style="color: blue;">●</span> Input neurons</li> <li><span style="color: red;">→</span> <math>X-h</math> weights</li> <li><span style="color: orange;">→</span> Recurrent weights</li> <li><span style="color: blue;">→</span> <math>H-Y</math> weights</li> <li><span style="color: grey;">●</span> Hidden neurons</li> <li><span style="color: red;">●</span> Output neurons</li> </ul>	<ul style="list-style-type: none"> <li>• Loops for storing information</li> <li>• Handles sequence data</li> </ul> Pros: <ul style="list-style-type: none"> <li>• Remembers previous inputs</li> <li>• Variable-length support</li> </ul> Cons: <ul style="list-style-type: none"> <li>• Gradient issues</li> <li>• Longer training times</li> </ul>	
<b>c</b> SNN 	<ul style="list-style-type: none"> <li> Input spike trains</li> <li><span style="color: blue;">●</span> Input neurons</li> <li><span style="color: red;">→</span> Weights</li> <li><math>\Sigma</math> Summation</li> <li> Output LIF neuron</li> </ul>	<ul style="list-style-type: none"> <li>• Discrete spikes for communication</li> <li>• Suited for event-driven data</li> </ul> Pros: <ul style="list-style-type: none"> <li>• Energy-efficient</li> <li>• Natural temporal representation</li> </ul> Cons: <ul style="list-style-type: none"> <li>• More complex to model and train</li> <li>• Limited tools and frameworks</li> </ul>	

**Fig. 3 | Bio-inspired neural network taxonomy.** Three architectures: FCN/CNN (a) RNN (b) and SNN (c). Also shown are their building blocks, key features, and implementation examples.

conductivity, while holes remain trapped in the FG. The ultrathin tunneling dielectric, featuring a high energy barrier and low defect density, suppresses carrier recombination between the FG and channel, thereby ensuring stable retention of the injected charge. And applying a positive gate voltage recombines trapped holes, enabling reversible resetting of optical conductance states. Although gate voltage precisely controls stored charge density, high operating voltages raise energy consumption concerns due to increased leakage currents.

Furthermore, significant progress in emerging electrochemical random-access memory (ECRAM) technology has emerged. ECRAM utilizes gate voltage pulses to move ions (such as  $\text{Li}^+$ ) in and out of the electrochemical layer, altering the charge distribution in the semiconductor channel<sup>62</sup>. This mechanism allows for effective tuning of photoresponsivity (Fig. 2h). Charge trapping transistors (CTTs) can modulate the channel in either a volatile or non-volatile way by trapping photo-generated carriers, due to defects present at the semiconductor’s interface or in its bulk (Fig. 2i). Volatile modulation discharges quickly, leading to a rapid return of channel conductivity to its initial state. In contrast, non-volatile modulation retains charge effectively and is less susceptible to external disturbances.

In advanced configurations, current OEMs have been further designed with two or multiple gate terminals to enable brain-inspired functionalities. For example, applying dual gate voltages ( $+V_g$  and  $-V_g$ ) can dynamically reconfigure p-type and n-type regions within a single channel, forming a reconfigurable p-n homojunction (Fig. 2j). This design effectively mimics synaptic weight updates in neural networks. Furthermore, multi-terminal FETs can transduce various environmental stimuli (e.g., light, pressure, temperature) into electrical signals, enabling co-modulation by multiple inputs and facilitating the artificial synapses that directly interface with

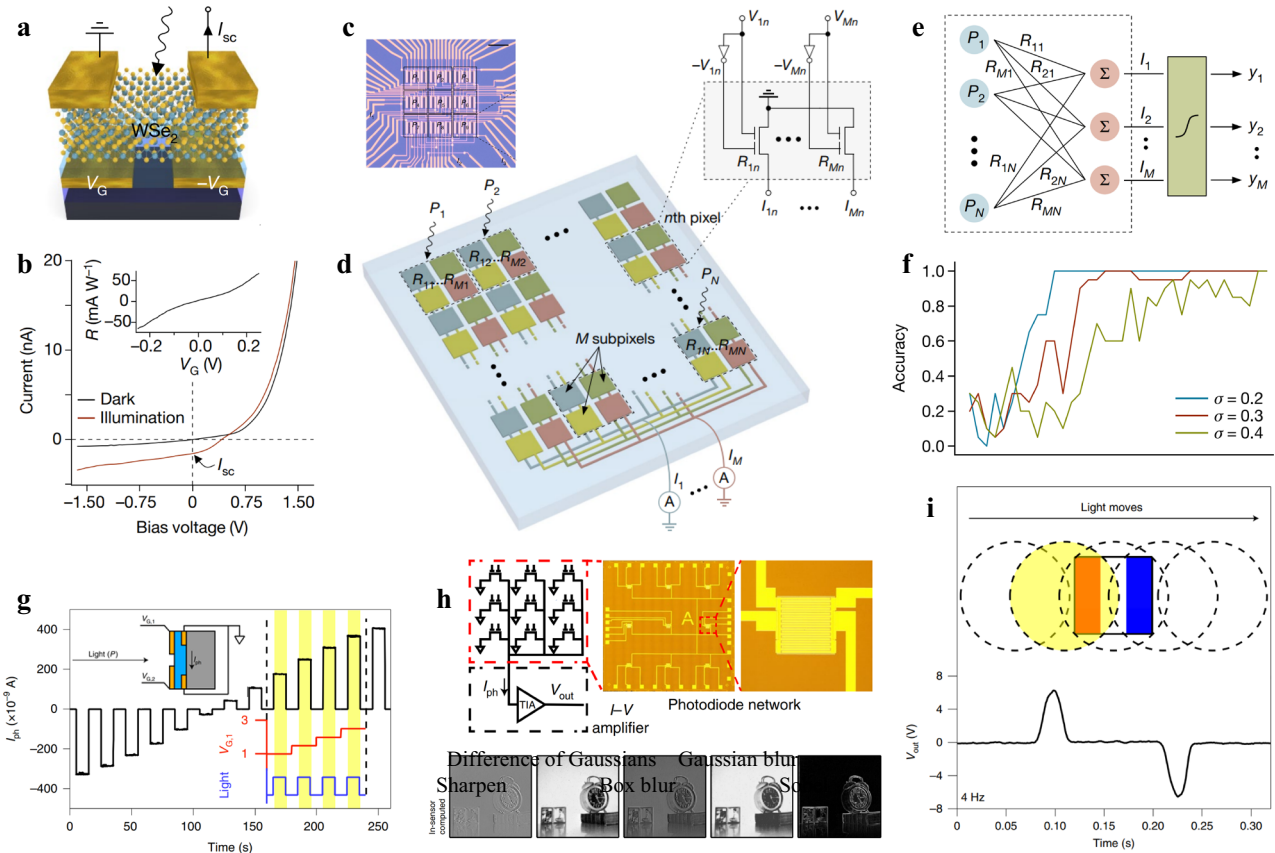
external environments. A representative example involves using coplanar multi-gate structures as modulating inputs, with the channel layer acting as a postsynaptic terminal. The system can achieve spatiotemporal coordinate and visual orientation recognition, resembling the hierarchical processing observed in the human visual cortex<sup>63</sup>. Additionally, multi-terminal FETs support parallel learning and dendritic integration by leveraging independent gate inputs to simulate synaptic summation and nonlinear activation. This architecture reduces the hardware complexity of neural networks by integrating multiple functions into a single device. However, research on neuromorphic multi-terminal FETs is still at an early stage.

### Bio-inspired neural networks

Integration of artificial synapses and neurons into artificial neural networks (ANNs) has been realized in various architectures including CNNs, RNNs, and SNNs. Particularly in the emerging field of in-sensor computing, researchers have developed novel implementations using optoelectronic devices: in-sensor FCNs/CNNs, in-sensor RNNs, and in-sensor SNNs (Fig. 3).

#### In-Sensor FCN/CNN

Recent research has focused on the construction of FCNs, particularly emphasizing single-layer perceptron networks at sensor terminals<sup>64</sup>. A key example is the introduction of reconfigurable sensor arrays capable of efficiently performing multiply-and-accumulate (MAC) operations, which are central for in-sensor processing pipelines (Fig. 3a). Consider a sensor array arranged in a  $M \times N$  configuration, where  $N$  indicates the number of pixels. Each pixel comprises  $M$  sensory elements (or subpixels), defined by the number of classification neurons. In this array, the photocurrent



**Fig. 4 | Optoelectronic devices for in-sensor FCNs/CNNs.** **a** Schematic of a single WSe<sub>2</sub> photodiode. **b** Current–voltage characteristic curve of a single WSe<sub>2</sub> photodiode, along with the gate voltage tunability of the photoresponsivity (inset). **c** Microscope image of the photodiode array consisting of 3 × 3 pixels, where each pixel comprises three WSe<sub>2</sub> photodiodes (subpixels). **d** Illustration of the photodiode array, with subpixels connected in parallel to generate output currents  $I_M$ . **e** Schematic of the classifier. **f** Accuracy of the classifier during training for varying artificial noise levels. Reproduced with permission<sup>64</sup>. Copyright 2020, Springer

Nature. **g** Measured photocurrent with pulsed light and stepped gate voltages ( $V_{g1}$ ,  $V_{g2}$ ). **h** Schematic and micrograph of a 3 × 3 photodiode network for analog MAC computation. The accumulated photocurrent is converted to voltage ( $V_{out}$ ) by a transimpedance amplifier (top). Various filtered images obtained with in-sensor computing (bottom). **i** Measured  $V_{out}$  of the photodiode array configured as an edge detection filter kernel, with the light spot moving across the array. Reproduced with permission<sup>114</sup>. Copyright 2022, Springer Nature.

generated by each sensor, represented by  $I_{mn} = R_{mn} \times P_n$ , illustrates the analog multiplication of the incident light power ( $P_n$ ) at the  $n$ -th pixel with the photoresponsivity ( $R_{mn}$ ) of the subpixel. The resulting currents are then summed along  $m$ -th interconnected sensor elements utilizing Kirchhoff’s current law, yielding  $M$  types of output photocurrent, expressed as  $I_m = \sum_{n=1}^N I_{mn} = \sum_{n=1}^N R_{mn} P_n$ . This operation can be succinctly articulated as the matrix–vector multiplication of  $\mathbf{I} = \mathbf{R}\mathbf{P}$ . Here,  $\mathbf{P} = (P_1, P_2, \dots, P_N)^T$  is the stimuli vector, and  $\mathbf{R} = (R_{mn})$  is the photoresponsivity matrix of the sensor array.

The matrix  $\mathbf{R}$  is reconfigurable through methods like electrostatic doping. By applying dual gate voltages of  $+V_g$  and  $-V_g$ , real-time shifts of the channel’s energy bands are achieved (Fig. 4a). This facilitates the formation of p-type and n-type regions within a single channel, effectively forming a p-n homojunction<sup>65</sup>. By varying the magnitude and polarity of the gate voltages, each sensor’s photocurrent or photoresponsivity can be adjusted to specific positive or negative values for a given incident light power (Fig. 4b). This effectively emulates changes in synaptic weight during training in neural networks, where “negative” refers to the direction (polarity) of the resulting photocurrent. Mueller et al. configured a 3 × 3 imaging array ( $N = 9$ ) with three photodiodes per pixel ( $M = 3$ ) (Fig. 4c) to achieve ultrafast image recognition at the ns level (Fig. 4f). Each photodiode employed WSe<sub>2</sub>, a 2D ambipolar semiconductor, as the channel material, with its photoresponsivity adjusted with dual gates through electrostatic doping. Additionally, the short-circuit photocurrents ( $I_{sc}$ ) produced by each photodiode

under optical illumination were summed according to Kirchhoff’s law by hard-wiring the devices in parallel (Fig. 4d).

Following this, we further discuss the implementation of in-sensor CNNs, utilizing a 3 × 3 convolutional kernel. The 3 × 3 sensor network is designed as an optoelectronic convolutional processor that employs a 3 × 3 map of photoresponsivity as an image filter kernel. To process an image, a 3 × 3 patch defined by light spots of varying power is projected onto the sensor network. Each sensor generates photocurrent given by  $I_k = R_k \times P_k$  (where  $k = 1-9$ ), which illustrates the multiplication of incident light power by photoresponsivity. The total photocurrent is accumulated using Kirchhoff’s current law, providing an outcome of the optoelectronic convolution. This output can be succinctly expressed as a dot product between the 1 × 9 incident light power vector and the 9 × 1 photoresponsivity vector. By sliding the 3 × 3 patch through the  $n \times n$  image and repeatedly performing the optoelectronic convolution, this generates a  $(n-2) \times (n-2)$  matrix of accumulated photocurrents, totaling  $(n-2)^2$  values, which represents the image processed with the convolutional filter. Donhee Ham et al. developed a dual-gate silicon p–i–n photodiode array that adjusts photoresponsivity through electrostatic doping for in-sensor optoelectronic convolutional image processing (Fig. 4g). This array utilizes a 3×3 photodiode network to implement seven different convolutional filters (Fig. 4h, i). However, these optoelectronic devices still suffer from the volatile nature of their photoresponsivities, which require a sustained gate bias to maintain.

### Non-volatile In-Sensor FCN/CNN

Integrating memory into in-sensor computing improves the versatility of visual systems, particularly for high-level computational tasks that require frequent data reading and writing. In this framework, “memory” refers to the system’s ability to store weights (photoresponsivity  $R$ ), rather than the photocurrents ( $I$ ). This is similar to that for in-memory computing ( $I_{\text{output}} = G \times V$ ), where weight conductance  $G$  is stored instead of input voltage ( $V$ ). The ability for non-volatile weight storage, which maintains a stable and linear photoresponsivity, facilitates the analog matrix–vector multiplication of  $I = RP$ . Thus, images can be directly recognized through the integrated visual systems without requiring external memory or processing units, achieving energy efficiency.

In early implementations, the weights of visual systems are stored in external memory and supplied to each sensor via cabling, which inevitably consume additional energy. More importantly, extremely complex circuits are required for the individual control of each unit in neural networks, which is almost impossible and unacceptable for large-scale sensory networks. For scalability, it is essential to store these weights locally. Table 1 presents some typical works that utilizes various mechanisms for local weight storage. For example, Wang et al. have developed ferroelectric-defined reconfigurable 2D  $\text{MoTe}_2$  photodiode arrays for in-memory sensing and computing, using organic ferroelectric P(VDF-TrFE) films as gate dielectrics (Fig. 5b, d)<sup>57</sup>. The weight is stored and programmed by gradually switching the ferroelectric domains, enabling 51 distinguishable weight states with linear and symmetrical characteristics (Fig. 5f). Once set, each sensor retains its photoresponsivity value and outputs a photocurrent with adjustable sign/magnitude. However, the claimed “memory” ability of some current neuromorphic sensors primarily refers to the storage of photocurrent, which can be used for some signal pre-processing operations.

Ionic migration offers an alternative mechanism for achieving reconfigurable and non-volatile photoresponsivity. By inducing an unbalanced ion distribution, it creates variations in electric potential and internal electric fields. Hu et al. demonstrated this concept by introducing a high concentration of sulfur vacancies (positive ions) at the metal-semiconductor (M-S) junctions of a M-MoS<sub>2</sub>-M structure (Fig. 5a)<sup>54</sup>. The migration of sulfur vacancies, driven by an in-plane electric field (voltage pulses), redistributes the local sulfur vacancy concentration in the MoS<sub>2</sub> channel, thus adjusting the Schottky barrier at the junctions. This modulation breaks the symmetry of the electric potential in M-S-M devices, affecting both the amplitude and polarity of the short-circuit photocurrent. Furthermore, like memristors, these photocurrent states are non-volatile as the barrier and vacancy concentration remain stable without an electric field. As a complementary strategy, Zhang et al. proposed an ion-induced localized-field modulation using core-sheath single-walled carbon nanotube@graphdiyne (SWNT@GDY) (Fig. 5c). This GDY structure features large, evenly distributed pores, allowing for high storage capacity of Li<sup>+</sup> ion. A small gate bias facilitates the intercalation of Li<sup>+</sup> ions from the electrolyte into GDY, where they remain stored after the removal of the gate bias. Applying positive or negative gate pulses permits precise control over the captured Li<sup>+</sup> ions in GDY, enabling linear modulation of photoresponsivity<sup>66</sup>.

### In-Sensor RNN

In dynamic machine vision, accurately predicting future motion trajectories is crucial for applications like video analysis, robotic vision, and self-driving technology. This predictive capability relies on the effective integration of current and previous visual inputs, enabling systems to perform complex tasks like speed extraction, motion anticipation, and motion extrapolation<sup>67</sup>. RNNs are specially designed to process sequences of data, learning from both past and present inputs to better model temporal dependencies within the visual data (Fig. 3b). Recently, in-sensor reservoir computing (RC) has emerged as a promising paradigm within the RNN framework<sup>68</sup>, significantly lowering computational costs while maintaining high efficiency, as only the readout layer requires training. In RC, a dynamic reservoir (an untrained recurrent layer) nonlinearly projects input data into a high-dimensional feature space. This transformation converts complex inputs

into linearly separable states, which are then combined through weighted summation during the training of the readout layer<sup>69,70</sup>.

Volatile optoelectronic devices exhibit intrinsic short-term memory capabilities, resulting in temporal evolution and summation of conductance under sequential light pulse stimulation (Fig. 6a). This behavior is ideal for dynamic processing and memory in hardware implementations of in-sensor RC<sup>71</sup>. Table 2 lists various volatile optoelectronic devices used in these applications. For example, the interaction between photo-generated carriers and charge trap centers slows the photocurrent response, allowing these devices to serve effectively as the reservoir layer of the RC system. As demonstrated by van Dijken et al., in ITO/ZnO/NSTO photomemristors, the trapping and de-trapping of photo-generated carriers by defect states (oxygen vacancies) at the ZnO/NSTO interface triggers a dynamic memristive response. Similarly, Chai et al. showed that in MoS<sub>2</sub> phototransistors, the photo-generated carriers trapped in the gap states of MoS<sub>2</sub> result in a sublinear increase in conductance over time, contributing to the device’s volatile characteristics.

Current studies are exploring novel device structures and systems to improve in-sensor RC capabilities. Adjustable temporal summation at sensory terminals allows for effective processing of motion perception across various scenarios (Fig. 6d). However, RC systems utilizing two-terminal volatile devices often face limitations with non-adjustable time-scales and relatively fixed nonlinear transformation functions. As previously mentioned, three-terminal phototransistors with an additional terminal for gate control offer a promising solution. They allow precise control of photo-generated carriers and a dynamic response to varying sensory stimuli. Research by Chai et al. has shown that MoS<sub>2</sub> phototransistor arrays can directly perceive different types of motion, ranging from slow (e.g., snail) to fast (e.g., cheetah). By varying gate voltages, they achieved a temporal resolution of 10 to 10<sup>6</sup> ms (Fig. 6b, c), effectively modulating the charge dynamics of shallow trapping centers. Additionally, there is increasing interest in developing energy-efficient integrated systems. Long et al. constructed an integrated photoelectronic RC system using volatile photoelectronic devices (amorphous GaO<sub>x</sub>) as the reservoir layer, connected to a memristor array (TaO<sub>x</sub>/HfO<sub>y</sub>) for latent fingerprint recognition (Fig. 6e, f). Remarkably, the system achieves over 90% recognition accuracy even under a 15% stochastic noise level.

### In-sensor SNN

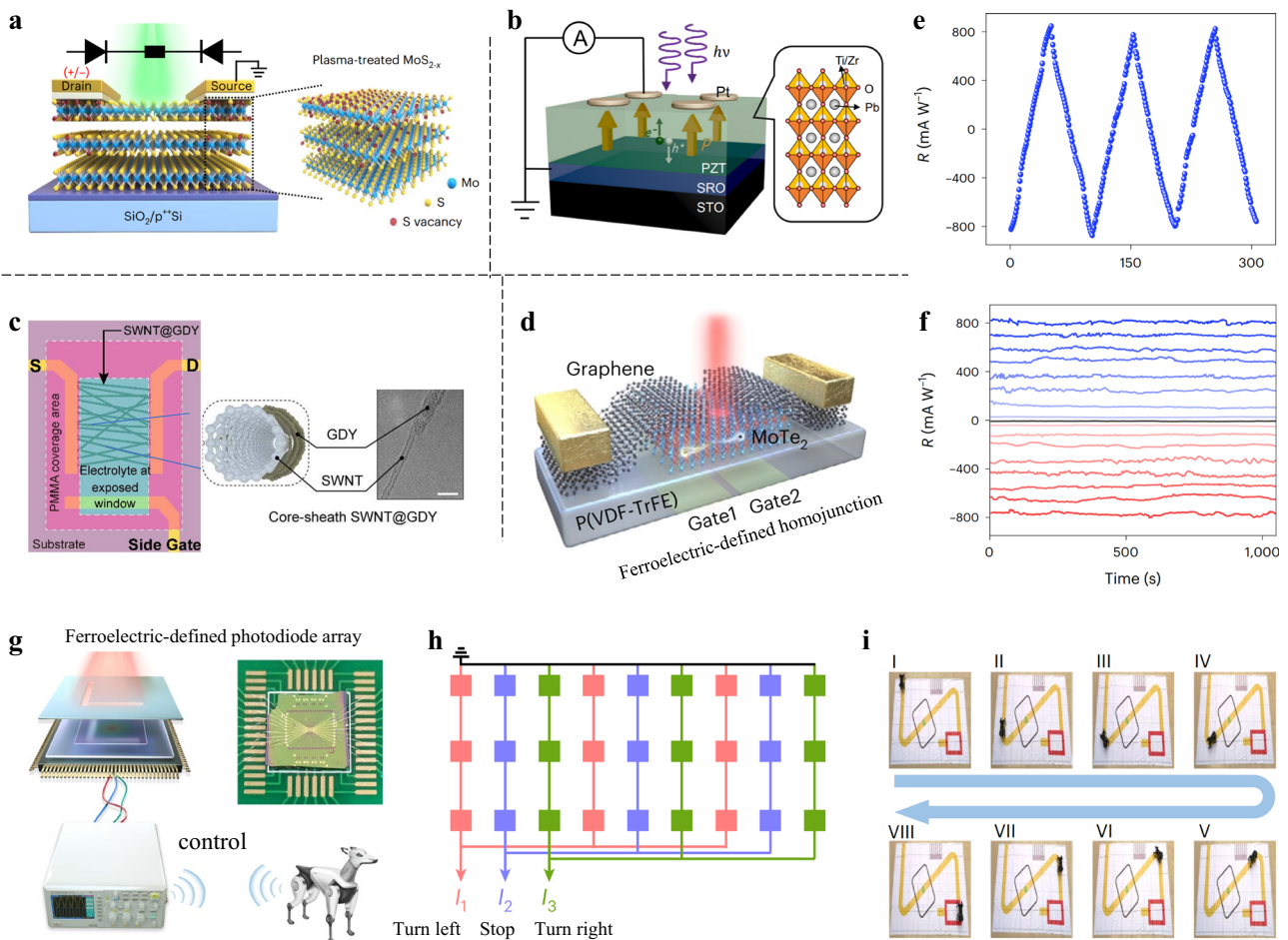
SNNs enable information processing in a way that closely resembles biological processes, thereby enhancing efficiency in tasks such as motion recognition. This network utilizes spiking neurons, primarily of the “LIF” type, to communicate through discrete spikes<sup>72</sup>. Each neuron receives input spikes (“events”) from a small area of visual space, altering its internal state (membrane potential), and producing an output spike (action potentials) when its state surpasses a predetermined threshold<sup>73</sup>. The computation performed by SNNs using LIF neurons relies primarily on simple operations such as the integration of input spikes through addition and the comparison of the internal state against a threshold value<sup>74</sup>. As a result, SNNs have significantly lower computational demands compared to traditional ANNs, which often depend on complex matrix operations involving extensive multiplications and additions<sup>75,76</sup>.

The optimal scenario for SNNs is to receive spike trains containing timing information from the ever-changing real world, just like how our brains operate (Fig. 3c). Currently, the performance of SNN training algorithms is evaluated using static image datasets that are converted into spike trains through various encoding techniques (Fig. 7a), such as rate coding, rate/temporal fusion (RTF) coding, or time-to-first-spike (TTFS) coding<sup>42,77,78</sup>. However, it is crucial to move beyond static benchmark datasets. More research is necessary to compile appropriate benchmark datasets from dynamic vision sensors (DVS) that can capture only dynamic changes in a scene<sup>79,80</sup>. Such dynamic sensory data can be combined with the temporal processing capability of SNNs to enable extremely low-power computing<sup>81</sup>.

**Table 1 | Summary of non-volatile optoelectronic devices for in-sensor FCN/CNN**

Device	Structure	Mechanism	Read voltage	Response speed	Light source	Energy consumption	Endurance	Multilevel states	Retention	Physical functionality	Array scale	Network	Vision task	Accuracy	Ref.
Pu/PZT/SRO (vertical)	2-terminal	Ferroelectric polarization	0 V	~ns	365 nm	3.1 nJ/programming; 0 J/inference	10 <sup>6</sup>	25	>18 days	Reconfigurable responsivities	1 × 9	SLP	Binary classification	100%	49
Gr/MoTe <sub>2</sub> /P(VDF-TrFE)/dual gates	4-terminal	Ferroelectric polarization	0 V	~µs	340-1310 nm	~0.1 pJ/programming; 0 J/inference	10 <sup>6</sup>	51	>24 h	Reconfigurable responsivities	3 × 9	SLP	Multi-class classification	Real-time control	57
Au/DIPAB/CuPc/substrate	2-terminal	Ferroelectric polarization	0 V	N/A	White-light LED	1.87 nJ/programming; 0 J/inference	N/A	N/A	>10 h	Reconfigurable responsivities	5 × 5	SLP	Letter classification	100%	95
Gr/Cr/PS/Gr (vertical)	2-terminal	Cu <sup>+</sup> ions	0 V	~ms	360 nm	0 J/inference	10 <sup>6</sup>	N/A	>3000 s	Reconfigurable responsivities	10 × 10	SLP	Letter classification	~ 76%	96
Au/Cr/SWNT@GDY/LiClO <sub>4</sub> /Cr/Au	3-terminal	Li <sup>+</sup> ions	0.1 V	~ms	400-1800 nm	50 aJ/programming	10 <sup>6</sup>	64	>10 <sup>4</sup> s	Reconfigurable responsivities	3 × 3 × 3	SLP	Classification	100%	66
Au/Cr/MoS <sub>2</sub> /SiO <sub>2</sub> /Si	2-terminal	Sulfur vacancies	0 V	N/A	520/830 nm	N/A	N/A	11	>1000 s	Reconfigurable responsivities	3 × 3	SLP	Classification	~90%	54
Gr/MoS <sub>2-x</sub> O <sub>y</sub> /SiO <sub>2</sub> /Si	2-terminal	Oxygen vacancies	0 V	~ms	Solar simulator	0 J/inference	10 <sup>2</sup>	7	>3000 s	Reconfigurable responsivities	3 × 3	SLP	Multi-class classification	96.44%	97
Au/BP/AHA/ITO	3-terminal	Floating gate	50 mV	N/A	1.5-3.1 µm	N/A	N/A	36	>2000 s	Reconfigurable responsivities	4 × 3	CNN	Digit classification	92%	98
Au/Cr/BP/MoS <sub>2</sub> /BN/Gr	3-terminal	Semi-floating gate	0.1 V	~µs	3900 (Blackbody)	1.8 fJ/programming	10 <sup>4</sup>	12	>1000 s	Reconfigurable responsivities	N/A	Faster-RCNN	Object detection	~ 90%	99

SLP single-layer perception, Gr Graphite, CrPS CuInP<sub>2</sub>S<sub>6</sub>, CuPc Copper phthalocyanine, DIPAB Diisopropylammonium bromide, AHA Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub>.



**Fig. 5 | Non-volatile optoelectronic devices for in-sensor FCN/CNN.** **a** Schematic of the M-MoS<sub>2</sub>-M device structure with back-to-back M-MoS<sub>2</sub> Schottky diodes. The MoS<sub>2</sub> under the source/drain electrodes is treated with O<sub>2</sub> and Ar plasma<sup>54</sup>. Copyright 2023, Springer Nature. **b** Schematics of the device structure of the Pt/PZT/SRO ferroelectric photosensor. Reproduced with permission<sup>49</sup>. Copyright 2022, Springer Nature. **c** Schematic of the device structure of core-sheath SWNT@GDY nanotubes based photosensor. Reproduced with permission<sup>66</sup>. Copyright 2023, AAAS. **d** Schematic of the ferroelectric-defined reconfigurable 2D homojunctions. **e** Symmetric and reversible modulation of the photoresponsivity tuned by

ferroelectric domains. **f** Long-term stability of positive states, negative states, and “zero” state. **g** Illustration of the hardware implementation process using a ferroelectric-defined photodiode array. The array recognizes the incoming pattern and sends a corresponding command to the robot dog via a wireless signal. **h** Patterns “L”, “T”, and “J” command the robot dog to move left, right and stop, respectively. Array 1 (*I*<sub>1</sub>), Array 2 (*I*<sub>2</sub>) and Array 3 (*I*<sub>3</sub>) are trained to identify patterns “L”, “T”, and “J”, respectively. **i** Robot dog navigation is determined by the corresponding real-time current profiles in (**h**). Reproduced with permission<sup>115</sup>. Copyright 2023, Springer Nature.

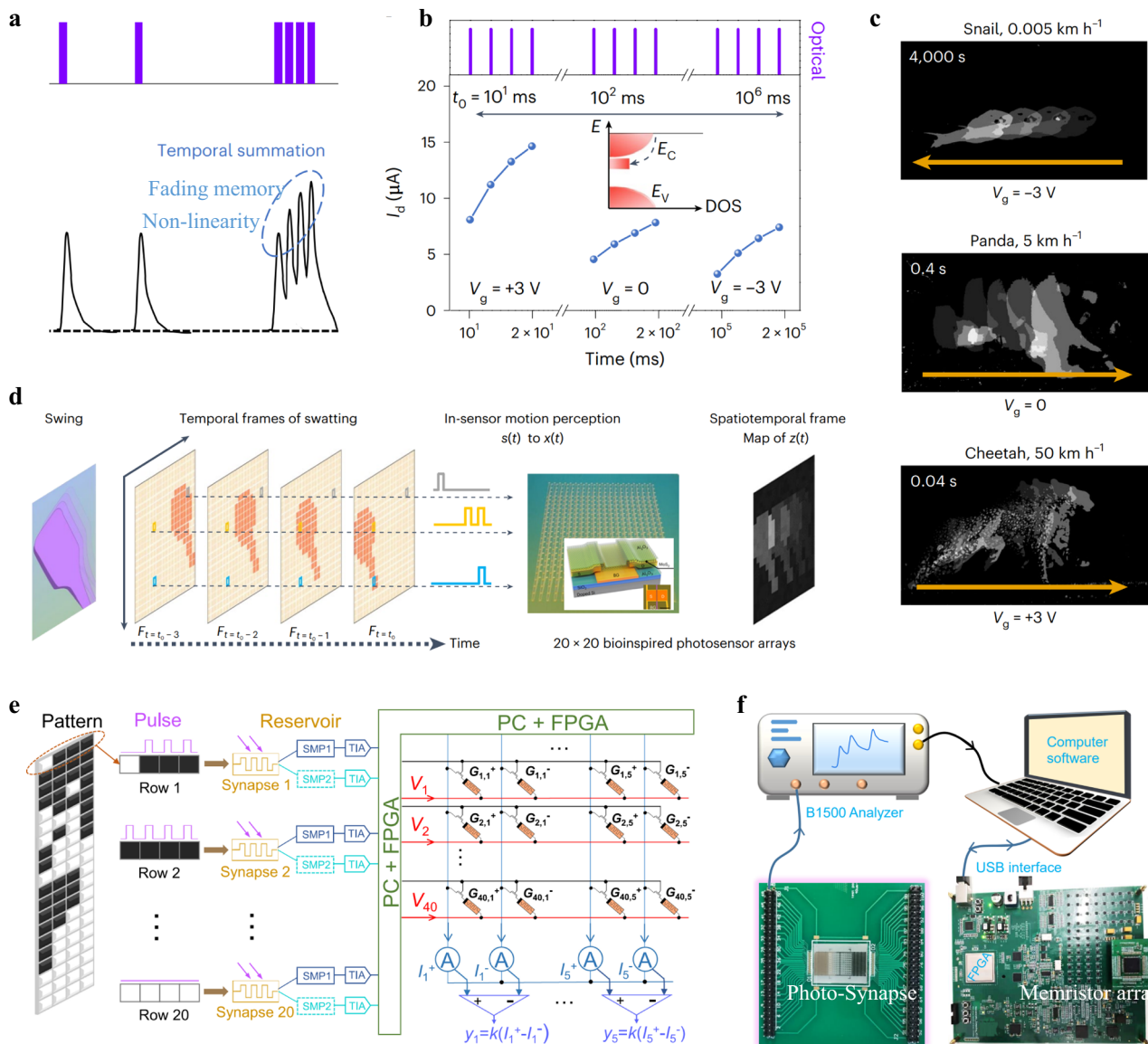
Furthermore, researchers have introduced the concept of in-sensor SNN. Unlike for static image recognition, in in-sensor SNNs, the MAC for motion recognition operates between the input light intensity change (vector) and the synaptic weight in the photodiode array (matrix), according to the physical law:  $\Delta I_m = \sum_{n=1}^N \Delta I_{mn} = \sum_{n=1}^N k \times R_{mn} \times \Delta P_n$ , where in which *k* is a constant, *R<sub>mn</sub>* is the photoresponsivity of the *m*-th subpixel in the *n*-th pixel, and  $\Delta P_n$  is the local light intensity change in the *n*-th pixel. The summed  $\Delta I_m$  is input to the output neurons with a LIF neuron. Chai et al. developed a computational event-driven vision sensor utilizing the 2D2R1C architecture (Fig. 7b), which converts dynamic motion into programmable, sparse, and informative spiking signals (Fig. 7e, f)<sup>82</sup>. Each pixel features two branches with opposite polarity and different photoresponse times, to generate event-driven spiking signals. They employ WSe<sub>2</sub> photodiodes with a floating dual-gate structure to maintain different non-volatile photoresponsivity (Fig. 7d). This allows for precise modulation of spiking signals amplitudes in each pixel, effectively emulating different synaptic weights in SNNs. The in-sensor SNNs can execute the motion recognition tasks with a temporal resolution of 5  $\mu$ s (Fig. 7c). Also, they designed a 2T2R1C cell with 2D MoS<sub>2</sub> phototransistors with an inherently high dynamic range of 140 dB<sup>83</sup>.

### Benchmarking metrics

Emerging optoelectronic devices must follow standardized key metrics for effective performance evaluation. Device-level metrics focus on the properties of individual component (e.g., energy efficiency, response speed, array scale, endurance (cycles)/retention). Table 3 summarizes the current state-of-the-art key metrics for both non-volatile/volatile OEMs for optoelectronic computing.

### Energy efficiency

The optimization of energy efficiency is critical for hardware implementations. As integration technology advances and computing density increases, the energy per operation becomes the defining metric for OEM performance. ECRAM devices based on ionic migration mechanisms demonstrate ultra-low programming energy (e.g., ~aJ/operation for ionic migration) and eliminate static power for state retention, while the reorientation of ferroelectric domains requires higher voltage and longer programming times compared to the drift of mobile cations in the electrolyte. In contrast, volatile OEMs (e.g., interfacial charge trapping) consume ~nJ/operation during dynamic operation, reflecting their reliance on continuous energy for temporal response. Remarkably,



**Fig. 6 | Optoelectronic devices for in-sensor RNN.** **a** Response characteristics of volatile MoS<sub>2</sub> phototransistors with nonlinear temporal summation characteristics when stimulated by light. **b** Processing four light stimulation pulses by MoS<sub>2</sub> phototransistors with time constants varying from 10 ms ( $V_g = +3$  V) to 10<sup>6</sup> ms ( $V_g = -3$  V). **c** Encoding spatiotemporal information with MoS<sub>2</sub> phototransistor at different time scales, tracking motion at speeds ranging from 0.005 to 50 km h<sup>-1</sup>. **d** Mapping spatiotemporal information with MoS<sub>2</sub> phototransistor arrays,

converting the temporal vision sequence  $s(t)$  into compressive states  $x(t)$ . Reproduced with permission<sup>105</sup>. Copyright 2023, Springer Nature. Schematic (e) and photograph (f) of the fully-hardware photoelectronic RC system, integrating a photoelectronic reservoir layer for feature outputs, a memristor array for training, and an FPGA controller. Reproduced with permission<sup>103</sup>. Copyright 2022, Springer Nature.

self-powered photovoltaic OEM architectures achieve zero-energy inference and energy use only during programming.

**Speed**

The response time of optoelectronic devices greatly impacts the working speed of vision sensors. It is determined by the materials and device structures. The charge-trapping-based devices inherently suffer from millisecond-scale latency due to the charge-trapping/de-trapping dynamics, creating a critical bottleneck for high-speed imaging. Remarkably, p-n junctions with an internal electric field can realize ultrafast photoresponse<sup>64</sup>. For volatile OEMs, their conductance dynamics are exploited for computing functions, such as short-term memory for motion recognition. Therefore, the temporal evolution of their conductance also determines the speed of processing the dynamic visual input.

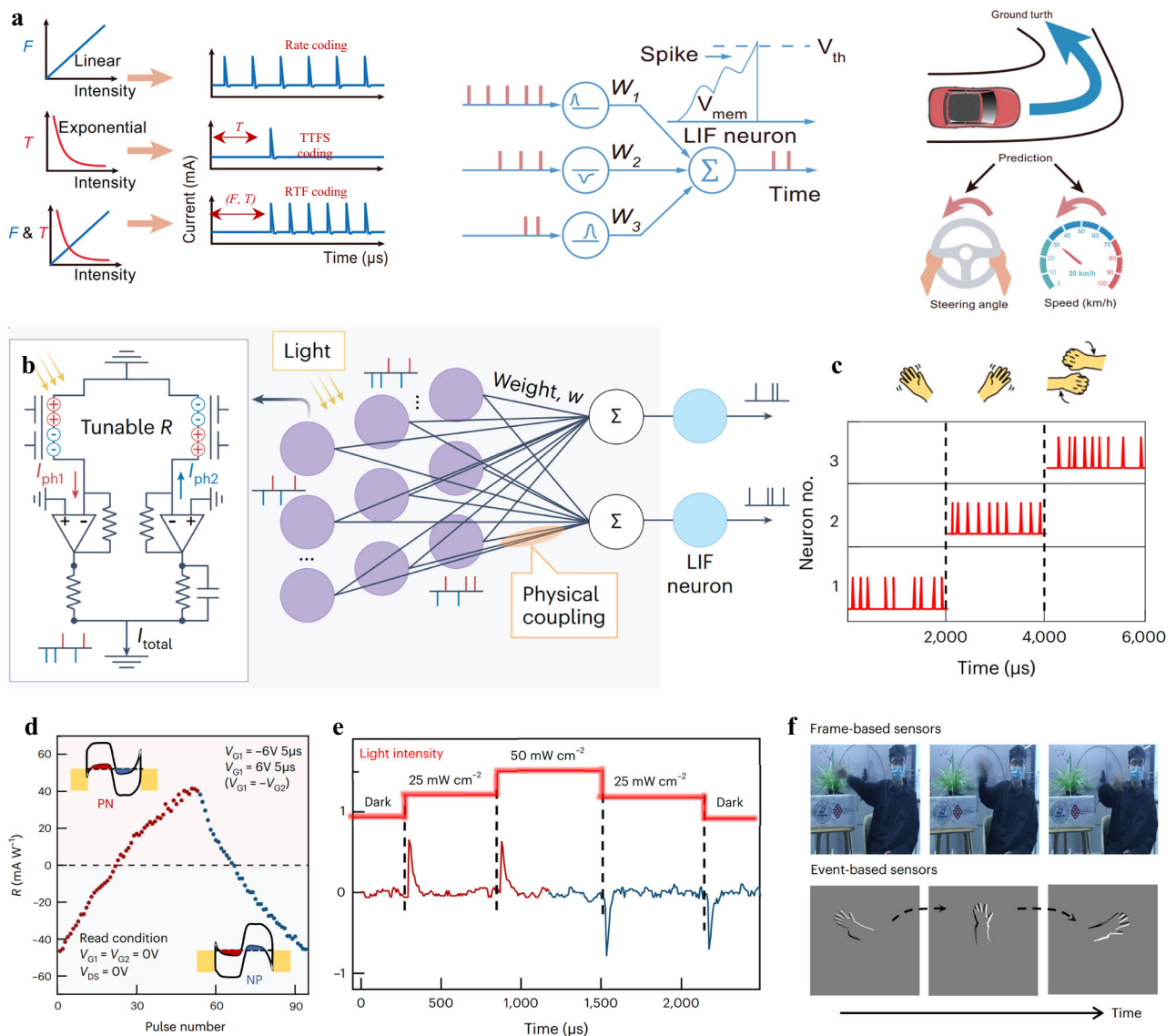
**Array scale**

As computing tasks become more complex, integrating more processing units on a single chip is now critical. Oxide semiconductors (e.g., ZnO) offer distinct advantages for scalable fabrication, including low device-to-device variation, low-temperature processing, and CMOS compatibility<sup>52</sup>. However, these materials primarily respond to violet light, limiting their utility in environmental sensing and high-energy imaging. Emerging materials like perovskites<sup>34</sup> demonstrate great potential for large-area fabrication and detection across visible to near-infrared wavelengths. Yet, they degrade rapidly when exposed to air, severely reducing their long-term stability. Recent advances in uniform large-area fabrication and material engineering aim to mitigate these limitations, paving the way for robust, high-performance optoelectronic systems.

**Table 2 | Summary of volatile optoelectronic devices for in-sensor RC**

Device	Structure	Mechanism	Light source	Energy per operation	Array scale	Physical functionality	Network	Vision task	Accuracy	Ref.
ITO/ZnO/NSTO (vertical)	2-terminal	Interfacial charge trapping	Blue LED	15 nJ	5 × 5	Dynamic temporal response	RC	Language learning	97.3%	100
Au/P(VDF/TFE)/Cs <sub>2</sub> AgBr <sub>2</sub> /ITO (vertical)	2-terminal	Photo-generated carriers pinning	445 nm	0	5 × 5	Dynamic temporal response	RC	Motion recognition Audio-to-motion prediction Face classification	100% ~90% 99.97%	101
Au/ZnO/Ni/GZO/TiN (vertical)	2-terminal	Charge trapping in ZnO/Ni	365 nm	N/A	N/A	Dynamic temporal response	RC	Dynamic vehicle flow detection Digit recognition	100% 90.45%	102
Pd/TiO <sub>2</sub> /ZnO/TiN (vertical)	2-terminal	Interfacial charge trapping/ conductive filaments in ZnO	405–650 nm	2 nJ	128 × 8	Non-volatile optoelectronic response	CNN	Human action classification Image pre-processing	97.14% 96.1%	52
Au/Ti/Amorphous GaO <sub>x</sub> /SiO <sub>2</sub> /Si	3-terminal	Oxygen-vacancy-related trapping	254 nm	0.5 nJ	20 × 4	Dynamic temporal response /Non-volatile optoelectronic response	Template-matching algorithm	Object tracking	96.1%	
Au/Cr/SnS/SiO <sub>2</sub> /Si	3-terminal	Sn/S vacancies	725 nm	85 nJ	1 × 5	Dynamic temporal response	RC	Human motion recognition Fingerprint recognition	91.2% ~ 90%	103
Au/Cr/MoS <sub>2</sub> /Al <sub>2</sub> O <sub>3</sub> /Cr/Au	3-terminal	Shallow trapping centers in MoS <sub>2</sub>	660 nm	N/A	20 × 20	Dynamic temporal response	RC	Short sentences of language classification Multiscale motion recognition	91% 99.2%	104 105
BTBT-C10/PSEBMA/ITO (vertical)	3-terminal	Interfacial charge trapping	White light	0.25 fJ	8 × 8	Dynamic temporal response	RC	Motion trajectory recognition	95.56%	106
Au/P3HT/UCNP-s@SiO <sub>2</sub> /SiO <sub>2</sub> /Si	3-terminal	Charge trapping	980 nm	N/A	1 × 8	Dynamic temporal response	RC	NIR static digit recognition NIR dynamic digit recognition	91.13% 90.07%	107
Au/p-NDI/SiO <sub>2</sub> /Si	3-terminal	Charge trapping	White light	N/A	N/A	Dynamic temporal response	RC	Nonlinear dynamic equation forecasting Multitask recognition	NMSE: 1.06 × 10 <sup>-3</sup> 88.0%	108
Au/Cr/MoS <sub>2</sub> /h-BN/Te/SiO <sub>2</sub> /Si	3-terminal	Charge trapping in MoS <sub>2</sub>	450/532/635 nm	N/A	N/A	Dynamic temporal response	RC	Dynamic hand gesture recognition Multimode recognition	98.62% 90.77%	109
Au/Pd/α-In <sub>2</sub> Se <sub>3</sub> /Al <sub>2</sub> O <sub>3</sub> /Si	3-terminal	Ferroelectric polarization	655 nm	N/A	N/A	Dynamic temporal response	RC	Multimode digit recognition Noisy QR code recognition MSO task forecasting	86.10% 98.60% NRMSE 0.105	110

NSTO Nb-doped SrTiO<sub>3</sub>, PSEBMA poly (sulfobetaine methacrylate), UCNPs upconversion nanoparticles, P3HT poly (3-hexylthiophene), IMO multiple superimposed oscillator, NRMSE normalized root-mean-square error, NMSE normalized mean squared error.



**Fig. 7 | Optoelectronic devices for in-sensor SNN.** **a** Greyscale images are encoded into series of spike trains by the rate coding, TTFS coding, and RTF coding schemes, and fed into a trained SNN to predict the vehicle speed and steering angle. Reproduced with permission<sup>77</sup>. Copyright 2024, Springer Nature. **b** An in-sensor SNN created by combining event-driven characteristics and in-sensor computing. Reproduced with permission<sup>116</sup>. Copyright 2023, Springer Nature. **c** Output spike situation in each neuron when the three types of motions are performed sequentially. **d** Non-volatile and programmable photoresponsivity under a series of gate voltage

pulses, proving its feasibility for use as the synaptic weights in neural networks. **e** Event-based sensors generate programmable spikes when light intensity changes. **f** Comparison between frame- and event-based vision sensors. The bottom panel shows the output signal after the pixel sensing process of event-based sensors, and only in the regions with the change of the light intensity changes generate positive or negative current spikes. Reproduced with permission<sup>82</sup>. Copyright 2023, Springer Nature.

**Endurance (cycles)/retention**

In non-volatile memory systems, endurance and retention are two critical operational metrics. ECRAM devices utilizing ionic migration mechanisms exhibit unstable properties in the programmed conductivity state, due to inherent material limitations. The migration of cations through liquid electrolytes induces spontaneous ion redistribution, leading to conductivity decay. In contrast, devices based on ferroelectric polarization mechanisms mainly exhibit much longer non-volatile retention performances of the programmed conductance state. Their stability derives from fundamental physical barriers that maintain polarization through coercive field thresholds. Reorienting the direction of polarized domains in ferroelectric materials is a passive process requiring an external bias programming signal. Therefore, the programmed conductance state can be maintained for a long time if there are no additional high-voltage programming attempts.

**Perception range**

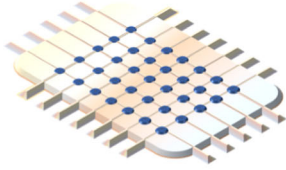
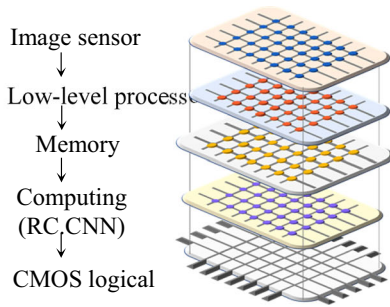
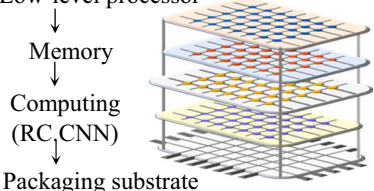
The perception range of optoelectronic devices refers to their ability to sense light intensity, speed, and wavelength. Adaptation in these devices enables the adjustment of the detection range, thus expanding their perception ability in different dimensions. Spatial adaptations allow optoelectronic devices to accurately capture details in both shadows and highlights within a single frame. Temporal adaptations broaden the scope of motion perception, extending it from static objects to high-speed movements. Additionally, spectrum adaptation fine-tunes the device’s sensitivity to various wavelengths, improving color detection under varying lighting conditions.

**Integration technologies**

The practical use of OEMs relies not just on diverse material systems and device structures but also on effective circuit design and processing technologies<sup>85,86</sup>. Traditional integration methods, particularly those used

**Table 3 | Summary of performance metrics of optoelectronic devices**

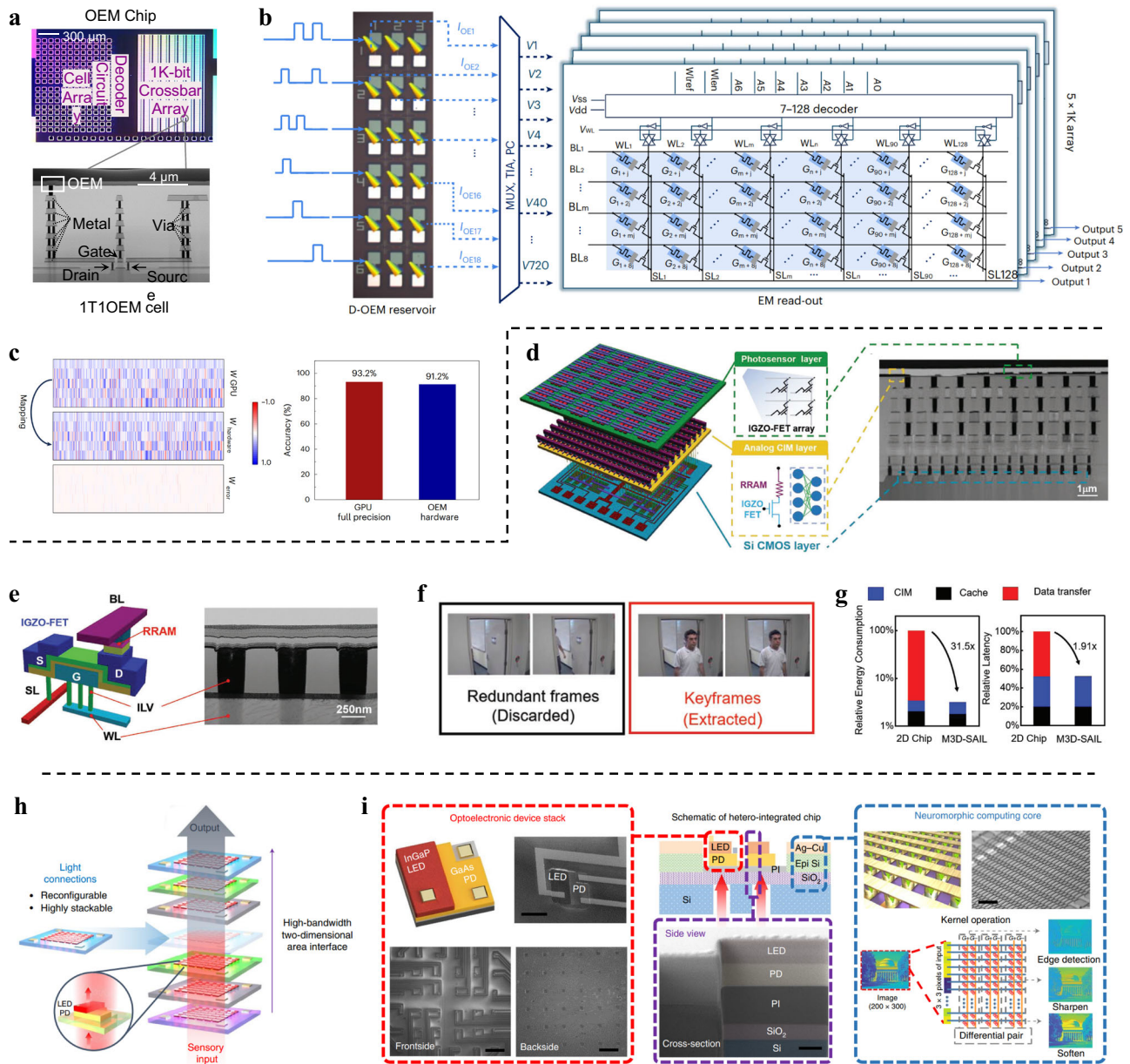
Device type	Device mechanism	Read voltage	Perception range	Response speed	Energy/operation	Array Scale	Endurance (Cycles)	State number	Retention	Physical functionality	Vision task
Non-volatile OEMs	Ferroelectric polarization	~0 V	UV-NIR	~ $\mu$ s-ns	~nJ/programming	Low	$>10^6$	$>51$	$>18$ days	Reconfigurable responsiveness	Feature extraction/Classification
	Ionic migration	$<0.1$ V	VIS-NIR	~ms	~aJ/programming	Low-medium	$\sim 10^5$	$>64$	$\sim 10^4$ s	Reconfigurable responsiveness	Feature extraction/Classification
	Charge trapping	~0 V	N/A	~ms	N/A	Low	$>10^2$	$>11$	$\sim 10^3$ s	Reconfigurable responsiveness	Feature extraction/Classification
	Floating gate	$<0.1$ V	N/A	~ $\mu$ s	N/A	Low	$>10^4$	$>12$	$\sim 10^3$ s	Reconfigurable responsiveness	Feature extraction/Classification
Volatile OEMs	Interfacial charge trapping	N/A	UV-VIS	N/A	~fJ-nJ	Low-medium	N/A	N/A	N/A	Dynamic temporal response	Static/Motion recognition
	Ferroelectric polarization	N/A	VIS	N/A	N/A	Low	N/A	N/A	N/A	Dynamic temporal response	Multiscale motion recognition

	<b>a</b>	<b>b</b>	<b>c</b>
Architecture	<p>Planar SoC</p> 	<p>Monolithic 3D Integration</p> <p>Image sensor ↓ Low-level processor ↓ Memory ↓ Computing (RC, CNN) ↓ CMOS logical</p> 	<p>Heterogeneous 3D Integration</p> <p>Image sensor ↓ Low-level processor ↓ Memory ↓ Computing (RC, CNN) ↓ Packaging substrate</p> 
Type	2D Flat, planar layout	3D, ILVs Stacked, vertical layout	3D, TSV Stacked, vertical layout
Key features	<p>Combines multiple functions on a single chip</p> <p>Pros:</p> <ul style="list-style-type: none"> <li>• Cost-effective</li> <li>• Simple and mature manufacturing</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Performance bottlenecks</li> <li>• Bandwidth constraints</li> </ul>	<p>Vertical stacking for reduced signal travel (10x bandwidth)</p> <p>Pros:</p> <ul style="list-style-type: none"> <li>• High-density</li> <li>• Low latency (<math>&lt;10</math> ns)</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Higher heat density</li> </ul>	<p>Flexible integration of sensors, processors, and memory in one package</p> <p>Pros:</p> <ul style="list-style-type: none"> <li>• High-density</li> <li>• Versatile</li> </ul> <p>Cons:</p> <ul style="list-style-type: none"> <li>• Manufacturing complexity</li> <li>• Higher heat density</li> </ul>

**Fig. 8 | Comparison of different integration technologies.** Three types of integration architectures: planar SoC (a) monolithic 3D integration (b) and heterogeneous 3D integration (c).

in printed circuit boards, often result in significant distances between sensing and processing units, which adversely impact integration density and efficiency. Thus, advancing integration technologies is essential for achieving low-power and low-latency architectures<sup>87,88</sup>. In-sensor computing architecture performs information processing within sensor networks, effectively eliminating the physical separation between sensing and computing units. Additionally, various integration technologies, including planar System on Chip (SoC), monolithic three-dimensional (3D) integration, heterogeneous 3D integration, and so on, are being actively explored.

Planar SoC integrates both sensing and processing units on a single substrate, typically arranging components in a flat, 2D layout. However, this integration technology limits the density and complexity compared to 3D integration, and longer interconnects can lead to increased signal latency due to extended data pathways between components. Despite these limitations, this design simplifies manufacturing and enabling better thermal management (Fig. 8a). Various visual sensor chips have utilized planar SoC technology for tasks such as feature extraction and high-level processing<sup>89</sup>. Wu et al. demonstrated a fully integrated 1-kb array with  $128 \times 8$  one-transistor-one-OEM cells integrated with Si

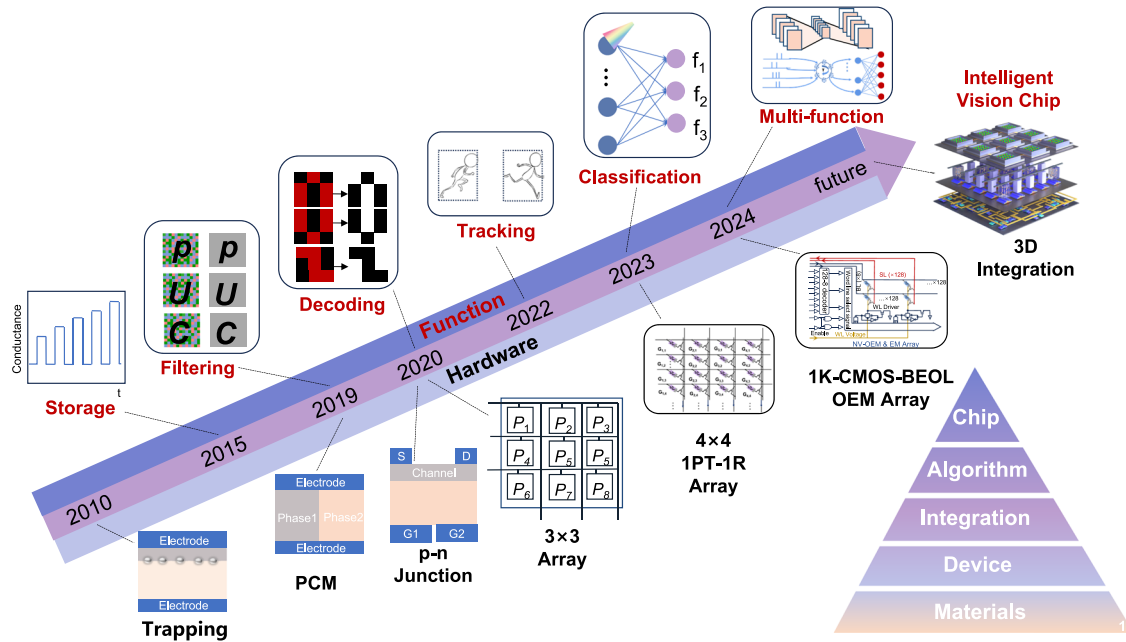


**Fig. 9 | Integration technologies for in-sensor computing.** **a** Micrograph of the  $128 \times 8$  crossbar array (top), and a cross-sectional high-resolution TEM image showing the 1T1OEM structure with a Pd/TiO<sub>x</sub>/ZnO/TiN stack on top of the drain terminal of the Si transistor (bottom). **b** Circuit diagram of a fully OEM-based in-sensor RC system. **c** Distributions of the  $(720 \times 5)$  target conductance weights, the OEM hardware weights and the error values (left). A comparison of the recognition accuracy after training for the GPU full precision simulation and the fully OEM-based in-sensor RC hardware (right). Reproduced with permission<sup>52</sup>. Copyright 2024, Springer Nature. **d** Schematic of the fabricated M3D-SAIL chip. **e** Schematic of the 1T1R cell with electrical connections to the Si CMOS circuits underneath.

**f** Schematic of video keyframe extraction task implemented by M3D-SAIL. **g** Energy consumption and latency benchmark of the M3D-SAIL chip. Reproduced with permission<sup>90</sup>. Copyright 2022, WILEY-VCH. **h** Schematic of stackable hetero-integrated chips with chip-to-chip communication in a sensor computing system. The first layer senses input and enables parallel processing over the 3D hetero-integrated chip stack. **i** Schematic of a hetero-integrated chip, mainly consisting of (left) an optoelectronic device stack with LED/PD arrays aligned with backside holes via deep reactive-ion etching, and (right) a neuromorphic computing core featuring a memristor array that performs three types of  $3 \times 3$  kernel operations. Reproduced with permission<sup>91</sup>. Copyright 2022, Springer Nature.

CMOS logic circuits (Fig. 9a). These Pd/TiO<sub>x</sub>/ZnO/TiN OEMs exhibit multi-modal characteristics by integrating functionalities of electronic memristors, dynamic OEMs, and non-volatile OEMs<sup>52</sup>. A fully OEM-based in-sensor RC system was thus developed for human motion recognition, where an 18-dynamic OEM array implements the reservoir layer, while five identical 1k electronic memristor arrays (mapping the  $720 \times 5$  weights using  $90 \times 8$  devices on each array) implement the readout layer (Fig. 9b). This monolithic integration of the multi-functional OEM with Si CMOS provides a cost-effective platform for various in-sensor computing applications (Fig. 9c).

The 3D integration enables the vertical stacking of different functional components, such as sensors, memory, computing, and communication units, into multiple tiers within a single system. This approach offers key advantages over planar geometry, including higher bandwidth for processing large amounts of data from sensor networks, a reduced footprint that saves space within the system, and shorter connection routing that minimizes delays for low-latency communication. Monolithic 3D integration adopts a microfabrication process with ultra-dense interlayer vias (ILVs) to build on-chip interconnects among functional layers. With a significantly smaller pitch than through-silicon vias (TSVs) used in heterogeneous 3D



**Fig. 10 | Development roadmap of in-sensor computing technology.** The development of this system encompassed fundamental materials to chip system integration, involving material science, physical mechanism, device design, fabrication, algorithm, and system integration. For the hardware, the research focused on the combination of different physical properties of single devices, gradually advancing

towards small-scale arrays and three-dimensional architectures. For the functionality, there has been a clear progression from low-level processing to high-level processing, such as storage, filtering, and denoising, and then moving on to more advanced tasks such as decoding, target tracking, and the capability to process complex visual scenes through a combination of multiple neural networks.

integration, ILVs enable higher integration density and ultrahigh on-chip data bandwidth (Fig. 8b). Wu et al. demonstrated the monolithic 3D integration of an IGZO-FET-based photosensor array for optical sensing, a one-transistor-one-resistor (1T1R) array using IGZO-FET and resistive random-access memory (RRAM) for analog computing-in-memory, and Si CMOS circuits for control logic (Fig. 9d, e), referred to as M3D-SAIL<sup>90</sup>. This approach exploits the high-bandwidth on-chip data transfer and parallel computing-in-memory cores for energy-efficient processing in video key-frame extraction (Fig. 9f and g), achieving 31.5× lower energy consumption and 1.91× faster computing speed compared to its 2D counterpart.

In heterogeneous 3D integration, copper microbumps and TSVs are commonly used to stack separately fabricated dies. While this technology increases area overhead, manufacturing complexities, and introduces parasitic capacitances that limit overall performance, it offers greater flexibility for specific applications. Unlike traditional monolithic 3D integration, which often relies on a single microfabrication process and may restrict design diversity, heterogeneous 3D integration allows for the combination of different semiconductor technologies within a single integrated package. To further enhance the capabilities of current heterogeneous 3D integration, Kim et al. introduced a reconfigurable integration technology that employs optoelectronic device stack including GaAs photodetectors (PDs) and InGaP light-emitting diodes (LEDs) for chip-to-chip communication in each heterogeneous chip layer, making it possible to stack and restack a variety of chip modules (Fig. 9h)<sup>91</sup>. Further in each stackable chip, neuromorphic computing processors embedded with a 32 × 32 Ag-Cu-alloy-based Si memristor crossbar array serves as deep learning hardware accelerators, efficiently processing analog data near photosensors (Fig. 9i). This system directly classifies information from a light-based image source. The initial data generated by the chip’s eye layer responds to external optical stimuli and is then processed through memristor-based hardware accelerators.

**Roadmap**

As illustrated in Fig. 10, it showcases the present research progress and future developments in the field of in-sensor computing. In the early stages,

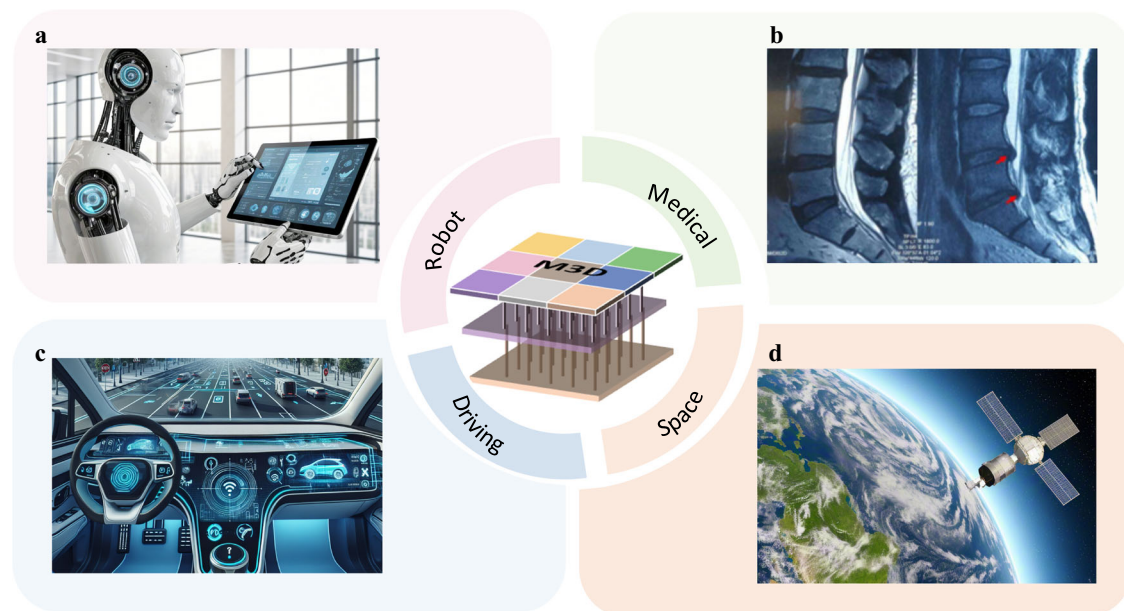
the primary focus was on the investigation of individual devices. These devices incorporated fundamental physical mechanisms such as trapping/detrapping<sup>92</sup>, photo-induced phase changes, p-n junctions, heterojunctions<sup>93</sup>, and chemical doping<sup>94</sup>, among others. The main goal of these foundational devices was to facilitate preliminary low-level pre-processing tasks such as analog storage<sup>50</sup>, image denoising<sup>17</sup>, and decoding. As the complexity and miniaturization of device structures advance, and with the maturation of array technology, the feasibility of validating more sophisticated visual processing tasks increases. For instance, CNN can be harnessed for cognitive functions, including but not limited to static object recognition and target tracking. Additionally, RC has the potential to execute motion recognition tasks. Looking forward, the development should focus on several key areas, including high-performance biomimetic devices, large-scale integration, multifunctional neural network algorithms, reducing system energy consumption, and expanding the application of intelligent terminal device chips:

**High-performance biomimetic devices**

To achieve the characteristics of artificial visual systems, such as high sensitivity, wide spectral response, high dynamic range, and adaptive adjustment, it is essential to research the physical mechanisms underlying the dynamic response of devices. Such investigations will drive breakthroughs in response sensitivity, wavelength range, response time, and biomimetic functional behavior. For instance, doping fabrication can be used to increase the density of states, expanding the response from ultraviolet to infrared wavelengths. Additionally, interface heterojunction design can enhance photocarrier separation efficiency while reducing device operating energy consumption. Furthermore, addressing engineering challenges is crucial to improving device uniformity and reliability, thereby establishing a solid hardware foundation for high-performance in-sensor computing systems.

**Large-scale, high-density integration**

It is crucial to develop integration technologies that align with the CMOS backend process and to embrace a three-dimensional integration architecture to boost data processing efficiency. By strategically assigning sensing,



**Fig. 11 | Application prospects for intelligent terminal devices. a** Embodied intelligence. **b** autonomous driving. **c** Medical field. **d** Space exploration.

computational processing, and logic control functions across different layers, and by amplifying the information transmission density through interlayer connections, we can optimize system functionality. Pre-processing can be carried out at the sensing layer, while a variety of neural network tasks can be managed at the computing-in-memory layer. Furthermore, the flow of signals can be controlled at the logic layer.

### Addressing reliability challenges

OEMs face critical reliability challenges, including cycle-to-cycle variations and device-to-device discrepancies that impede practical deployment. Key concerns involve material and interface defects (e.g., polarization fatigue in ferroelectric devices and stochastic filament formation in CF devices) and material degradation under repeated operation (e.g., electrolyte decomposition in ECRAM). To enhance reliability, strategies include material optimization (e.g., defect-passivated high-quality ferroelectrics like  $\text{HfZrO}_2$  and solid-state electrolytes like LiPON), structural innovations (3D array integration, nanoscale electrode designs for filament confinement). Advanced fabrication techniques (e.g., atomic layer deposition) and CMOS-compatible processes reduce variability. Collaborative co-design of materials, processes, and architectures is essential to overcome reliability bottlenecks in large-scale in-sensor computing systems.

### Multifunctional neural network algorithms

We need to develop low-level to high-level processing algorithms and handle both static and dynamic visual tasks, simulating complex information processing, and thus mimicking the collaborative work of the dorsal and ventral pathways in the visual system to accomplish tasks across different time sequences, low illumination and multi-target tasks.

### Improving system energy efficiency

Generally, the energy consumption of CMOS sensor processors consists of the combined energy usage of the sensor, memory, computing system, and peripheral transmission system. To further evaluate the energy efficiency of in-sensor computing systems, it is essential to consider the total energy consumption of the entire system required to complete the task. The total energy consumption includes the operational energy of the optoelectronic array and the energy consumption of peripheral circuits, which can be calculated as follows:

$$\text{Total Energy Consumption} = (\text{Number of computational operations}) \times (\text{Electrical operation energy per device} + \text{Optical operation energy per}$$

device) + Peripheral system energy consumption. To enhance the energy efficiency of the system, the following optimization strategies can be employed:

- (1) Optimization of algorithm architectural design: Reduce the number of computational operations required by the algorithm to lower the overall computation.
- (2) Optimization of device functionality: Improve device materials and structures to reduce the energy consumption of individual operations. For example, utilizing photovoltaic P-N junctions can reduce the external electric field required for photoelectric responses, thereby minimizing energy consumption.
- (3) Enhancement of system integration: Optimize the design of peripheral circuits to reduce the energy consumption of non-core components.

By implementing these multi-level optimization strategies, the energy efficiency of in-sensor computing systems can be significantly improved, providing robust hardware support for high-performance computing.

### Application prospects for intelligent terminal chip

Optoelectronic multifunctional devices provide an efficient and flexible hardware foundation for intelligent terminal devices, demonstrating broad application prospects as shown in Fig. 11. In the field of embodied intelligence, highly adaptive optoelectronic devices can maintain exceptional performance even under extreme lighting conditions. By utilizing dynamically tunable optoelectronic response devices, 3D augmented reality processing can be achieved, enabling robots to efficiently handle complex real-world problems. In autonomous driving, hardware systems equipped with high-speed SNN support can enable rapid processing of visual information and motion control, significantly reducing latency and enhancing processing speed and system safety. In the medical field, intelligent vision systems can efficiently classify and extract critical pathological information, greatly shortening diagnosis time and improving accuracy, thereby effectively assisting doctors in evaluating and managing medical conditions. In space exploration, by extending the multispectral response capabilities of devices, functions such as cosmic ray detection and extraterrestrial material analysis can be realized, providing advanced technological support for deep-space exploration.

Throughout this process, it is vital to fully utilize emerging technologies and methodologies to build more efficient, reliable in-sensor computing systems. This may involve exploring new materials, adopting advanced

manufacturing processes, and innovating in algorithms and architectures. Through these efforts, we aim to transition in-sensor computing technology from the lab to practical applications, thereby providing robust support for the development of intelligent systems.

## Conclusions

This review provides an overview of bio-inspired OEMs for energy-efficient in-sensor computing. By integrating bio-inspired neural networks—particularly, in-sensor FCNs/CNNs for static recognition, RNNs for motion tracking, and SNNs for event-driven processing—these OEMs mimic retinal-cortical vision via innovative pathway processing. Looking ahead, it is crucial to further explore new materials, advanced manufacturing processes, and innovate algorithms and architectures to develop robust, large-scale in-sensor computing systems. Additionally, building standardized benchmarking metrics and reducing overall system energy consumption will be vital for progress. Broadening the application of intelligent terminal chips will also play a key role in advancing intelligent systems. We aim to build more efficient and reliable in-sensor computing systems, ultimately bridging the gap between lab and practical applications.

Received: 29 November 2024; Accepted: 28 April 2025;

Published online: 03 July 2025

## References

- Wang, W. et al. Neuromorphic sensorimotor loop embodied by monolithically integrated, low-voltage, soft e-skin. *Science* **380**, 735–742 (2023).
- Chen, Y. et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* **623**, 48–57 (2023).
- Pei, J. et al. Towards artificial general intelligence with a hybrid Tianjic chip architecture. *Nature* **572**, 106–111 (2019).
- Schaller, R. R. Moore's law: past, present and future. *IEEE Spectr.* **34**, 52–59 (1997).
- Yao, P. et al. Fully hardware-implemented memristor convolutional neural network. *Nature* **577**, 641–646 (2020).
- Zhou, F. & Chai, Y. Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020).
- Schuman, C. D. et al. Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput. Sci.* **2**, 10–19 (2022).
- Indiveri, G. & Douglas, R. Neuromorphic vision sensors. *Science* **288**, 1189–1190 (2000).
- Marx, V. The big challenges of big data. *Nature* **498**, 255–260 (2013).
- Bays, P. M. & Husain, M. Dynamic shifts of limited working memory resources in human vision. *Science* **321**, 851–854 (2008).
- Ding, H., Smith, R. G., Poleg-Polsky, A., Diamond, J. S. & Briggman, K. L. Species-specific wiring for direction selectivity in the mammalian retina. *Nature* **535**, 105–110 (2016).
- Liao, F. et al. Bioinspired in-sensor visual adaptation for accurate perception. *Nat. Electron.* **5**, 84–91 (2022).
- Ouyang, B. et al. Bioinspired in-sensor spectral adaptation for perceiving spectrally distinctive features. *Nat. Electron.* **7**, 705–713 (2024).
- Li, L. et al. Adaptive machine vision with microsecond-level accurate perception beyond human retina. *Nat. Commun.* **15**, 6261 (2024).
- Jiang, X. et al. Multiframe-integrated, in-sensor computing using persistent photoconductivity. *J. Semicond.* **45**, 092401 (2024).
- Zhou, G. et al. Full hardware implementation of neuromorphic visual system based on multimodal optoelectronic resistive memory arrays for versatile image processing. *Nat. Commun.* **14**, 8489 (2023).
- Zhou, F. et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).
- Syed, G. S., Zhou, Y., Warner, J. & Bhaskaran, H. Atomically thin optomemristive feedback neurons. *Nat. Nanotechnol.* **18**, 1036–1043 (2023).
- He, Z. et al. An organic transistor with light intensity-dependent active photoadaptation. *Nat. Electron.* **4**, 522–529 (2021).
- Pang, X. et al. Non-volatile rippled-assisted optoelectronic array for all-day motion detection and recognition. *Nat. Commun.* **15**, 1613 (2024).
- Dang, Z. et al. Object motion detection enabled by reconfigurable neuromorphic vision sensor under ferroelectric modulation. *ACS Nano* **18**, 27727–27737 (2024).
- Gao, Q. et al. Physical origin of planar linear dichroism in van der Waals semiconductors using main group elements. *Sci. China Inf. Sci.* **67**, 222402 (2024).
- Zhang, F. et al. Bandgap engineering of Bilns Nanowire for wide-spectrum, high-responsivity, and polarimetric-sensitive detection. *Adv. Funct. Mater.* **33**, 2306077 (2023).
- Wang, X. et al. Polarizer-free polarimetric image sensor through anisotropic two-dimensional GeSe. *Sci. China Mater.* **64**, 1230–1237 (2021).
- Wu, W.-Q., Wang, C.-F., Han, S.-T. & Pan, C.-F. Recent advances in imaging devices: Image sensors and neuromorphic vision sensors. *Rare Met.* **43**, 5487–5515 (2024).
- Van Essen, D. C., Anderson, C. H. & Felleman, D. J. Information processing in the primate visual system: an integrated systems perspective. *Science* **255**, 419–423 (1992).
- Priebe, N. J. & Ferster, D. Inhibition, spike threshold, and stimulus selectivity in primary visual cortex. *Neuron* **57**, 072901 (2008).
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
- Tamietto, M. & Morrone, M. C. Visual plasticity: blindsight bridges anatomy and function in the visual system. *Curr. Biol.* **26**, R70–R73 (2016).
- Peng, Z. et al. Multifunctional human visual pathway-replicated hardware based on 2D materials. *Nat. Commun.* **15**, 8650 (2024).
- Baden, T., Euler, T. & Berens, P. Understanding the retinal basis of vision across species. *Nat. Rev. Neurosci.* **21**, 5–20 (2020).
- Solomon, S. G. & Lennie, P. The machinery of colour vision. *Nat. Rev. Neurosci.* **8**, 276–286 (2007).
- Chen, X., Wang, F., Fernandez, E. & Roelfsema, P. R. Shape perception via a high-channel-count neuroprosthesis in monkey visual cortex. *Science* **370**, 1191–1196 (2020).
- Zheng, Y., Jia, S., Yu, Z., Liu, J. K. & Huang, T. Unraveling neural coding of dynamic natural visual scenes via convolutional recurrent neural networks. *Patterns* **2**, 100350 (2021).
- Zhang, S. & Wang, Z. Artificial cat eye camera for objects detection against complex backgrounds and varied lighting. *J. Semicond.* **45**, 120401–120403 (2024).
- Ran, W., Wang, Z. & Shen, G. Artificial hawk-eye camera for foveated, tetrachromatic, and dynamic vision. *J. Semicond.* **45**, 090401–090403 (2024).
- Bhattacharya, K., Kumar, M. & Das, S. Spike-Afferent Visual Optoelectronic Nerve (SAVON) for Retinomorph Sensing Applications. *IEEE J. Electron Devices Soc.* <https://doi.org/10.1109/JEDS.2024.3492101> (2024).
- Shao, J. et al. Artificial sensory neurons and their applications. *J. Semicond.* **46**, 011606 (2025).
- Kim, Y. et al. Two-terminal neuromorphic devices for spiking neural networks: neurons, synapses, and array integration. *ACS Nano* **18**, 34531–34571 (2024).
- Chen, Y. et al. Revolutionizing neuromorphic computing with memristor-based artificial neurons. *J. Semicond.* **46**, 1–11 (2025).
- Ren, Q. et al. Optoelectronic devices for in-sensor computing. *Adv. Mater.* 2407476 (2024).

42. Wang, X. et al. Vertically integrated spiking cone photoreceptor arrays for color perception. *Nat. Commun.* **14**, 3444 (2023).
43. Choi, H. et al. Spiking neural network integrated with impact ionization field-effect transistor neuron and a ferroelectric field-effect transistor synapse. *Adv. Mater.* 2406970 (2024).
44. Pickett, M. D., Medeiros-Ribeiro, G. & Williams, R. S. A scalable neuron built with Mott memristors. *Nat. Mater.* **12**, 114–117 (2013).
45. Beck, M. E. et al. Spiking neurons from tunable Gaussian heterojunction transistors. *Nat. Commun.* **11**, 1565 (2020).
46. Xu, L. et al. Ultrasensitive dim-light neuromorphic vision sensing via momentum-conserved reconfigurable van der Waals heterostructure. *Nat. Commun.* **15**, 9011 (2024).
47. Ielmini, D. & Wong, H.-S. P. In-memory computing with resistive switching devices. *Nat. Electron.* **1**, 333–343 (2018).
48. Prezioso, M. et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature* **521**, 61–64 (2015).
49. Cui, B. et al. Ferroelectric photosensor network: an advanced hardware solution to real-time machine vision. *Nat. Commun.* **13**, 1707 (2022).
50. Tan, H. et al. An optoelectronic resistive switching memory with integrated demodulating and arithmetic functions. *Adv. Mater.* **27**, 2797–2803 (2015).
51. Hu, D.-C., Yang, R., Jiang, L. & Guo, X. Memristive synapses with photoelectric plasticity realized in ZnO<sub>1-x</sub>/AlO<sub>y</sub> heterojunction. *ACS Appl. Mater. Interfaces* **10**, 6463–6470 (2018).
52. Huang, H. et al. Fully integrated multi-mode optoelectronic memristor array for diversified in-sensor computing. *Nat. Nanotechnol.* **20**, 93–103 (2025).
53. Wang, T.-Y. et al. Reconfigurable optoelectronic memristor for in-sensor computing applications. *Nano Energy* **89**, 106291 (2021).
54. Li, T. et al. Reconfigurable, non-volatile neuromorphic photovoltaics. *Nat. Nanotechnol.* **18**, 1303–1310 (2023).
55. Fu, X. et al. Graphene/MoS<sub>2-x</sub>O<sub>x</sub>/graphene photomemristor with tunable non-volatile responsivities for neuromorphic vision processing. *Light Sci. Appl.* **12**, 39 (2023).
56. Song, Y. et al. Fatigue of ferroelectric field effect transistor: mechanisms and optimization strategies. *J. Semicond.* **46**, 1–14 (2025).
57. Tong, L. et al. 2D materials-based homogeneous transistor-memory architecture for neuromorphic hardware. *Science* **373**, 1353–1358 (2021).
58. Kumar, M. & Suri, M. Emerging non-volatile memory (NVM) technologies based nano-oscillators: Materials to applications. *J. Appl. Phys.* **136**, 160701 (2024).
59. Uk Lee, D., Jun Lee, H., Kyu Kim, E., You, H.-W. & Cho, W.-J. Low operation voltage and high thermal stability of a WSi<sub>2</sub> nanocrystal memory device using an Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> tunnel layer. *Appl. Phys. Lett.* **100**, 072901 (2012).
60. Suh, D. C. et al. Improved thermal stability of Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> high-k gate dielectric stack on GaAs. *Appl. Phys. Lett.* **96**, 142112 (2010).
61. Wang, Y. et al. Photonic synapses based on inorganic perovskite quantum dots for neuromorphic computing. *Adv. Mater.* **30**, 1802883 (2018).
62. Huang, H. et al. Electrolyte-gated transistors for neuromorphic applications. *J. Semicond.* **42**, 013103 (2021).
63. Xie, D. et al. Coplanar multigate MoS<sub>2</sub> electric-double-layer transistors for neuromorphic visual recognition. *ACS Appl. Mater. Interfaces* **10**, 25943–25948 (2018).
64. Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
65. Pan, C. et al. Reconfigurable logic and neuromorphic circuits based on electrically tunable two-dimensional homojunctions. *Nat. Electron.* **3**, 383–390 (2020).
66. Zhang, G.-X. et al. Broadband sensory networks with locally stored responsivities for neuromorphic machine vision. *Sci. Adv.* **9**, eadi5104 (2023).
67. Moon, J. et al. Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat. Electron.* **2**, 480–487 (2019).
68. Lukoševičius, M. & Jaeger, H. Reservoir computing approaches to recurrent neural network training. *Comput. Sci. Rev.* **3**, 127–149 (2009).
69. Liang, X. et al. Physical reservoir computing with emerging electronics. *Nat. Electron.* **7**, 193–206 (2024).
70. Cucchi, M., Abreu, S., Ciccone, G., Brunner, D. & Kleemann, H. Hands-on reservoir computing: a tutorial for practical implementation. *Neuromorph. Comput. Eng.* **2**, 032002 (2022).
71. Kim, J. et al. Analog reservoir computing via ferroelectric mixed phase boundary transistors. *Nat. Commun.* **15**, 9147 (2024).
72. Gallego, G. et al. Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154–180 (2020).
73. Roy, K., Jaiswal, A. & Panda, P. Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**, 607–617 (2019).
74. Davies, M. et al. Advancing neuromorphic computing with Loihi: A survey of results and outlook. *Proc. IEEE* **109**, 911–934 (2021).
75. Pfeiffer, M. & Pfeil, T. Deep learning with spiking neurons: opportunities and challenges. *Front. Neurosci.* **12**, 774 (2018).
76. Zhang, W. et al. Neuro-inspired computing chips. *Nat. Electron.* **3**, 371–382 (2020).
77. Li, F. et al. An artificial visual neuron with multiplexed rate and time-to-first-spike coding. *Nat. Commun.* **15**, 3689 (2024).
78. Wang, F. et al. A two-dimensional mid-infrared optoelectronic retina enabling simultaneous perception and encoding. *Nat. Commun.* **14**, 1938 (2023).
79. Wu, Y. et al. CMOS-compatible retinomorph Si photodetector for motion detection. *Sci. China Inf. Sci.* **66**, 162401 (2023).
80. Wu, S. E. et al. Retinomorph motion detector fabricated with organic infrared semiconductors. *Adv. Sci.* **10**, 2304688 (2023).
81. Liao, F., Zhou, F. & Chai, Y. Neuromorphic vision sensors: Principle, progress and perspectives. *J. Semicond.* **42**, 013105 (2021).
82. Zhou, Y. et al. Computational event-driven vision sensors for in-sensor spiking neural networks. *Nat. Electron.* **6**, 870–878 (2023).
83. Zhou, Y. et al. A 2T2R1C vision cell with 140 dB dynamic range and event-driven characteristics for in-sensor spiking neural network. In *International Electron Devices Meeting* 31–34 (IEEE, 2022).
84. He, Z. et al. Perovskite retinomorph image sensor for embodied intelligent vision. *Sci. Adv.* **11**, eads2834 (2025).
85. Lee, D. et al. In-sensor image memorization and encoding via optical neurons for bio-stimulus domain reduction toward visual cognitive processing. *Nat. Commun.* **13**, 5223 (2022).
86. Dang, B. et al. One-phototransistor-one-memristor array with high-linearity light-tunable weight for optic neuromorphic computing. *Adv. Mater.* **35**, 2204844 (2023).
87. Pendurthi, R. et al. Monolithic three-dimensional integration of complementary two-dimensional field-effect transistors. *Nat. Nanotechnol.* **19**, 970–977 (2024).
88. Jayachandran, D., Sakib, N. U. & Das, S. 3D integration of 2D electronics. *Nat. Rev. Electr. Eng.* **1**, 300–316 (2024).
89. Wang, S. et al. An organic electrochemical transistor for multi-modal sensing, memory and processing. *Nat. Electron.* **6**, 281–291 (2023).
90. Du, Y. et al. Monolithic 3D integration of analog RRAM-based computing-in-memory and sensor for energy-efficient near-sensor computing. *Adv. Mater.* **36**, 2302658 (2024).
91. Choi, C. et al. Reconfigurable heterogeneous integration using stackable chips with embedded artificial intelligence. *Nat. Electron.* **5**, 386–393 (2022).

92. Agnus, G. et al. Two-terminal carbon nanotube programmable devices for adaptive architectures. *Adv. Mater.* **22**, 702–706 (2010).
93. Zhang, Z. et al. All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition. *Nat. Nanotechnol.* **17**, 27–32 (2022).
94. Chen, K. et al. Organic optoelectronic synapse based on photon-modulated electrochemical doping. *Nat. Photonics* **17**, 629–637 (2023).
95. Cai, Y. et al. In-situ artificial retina with all-in-one reconfigurable photomemristor networks. *npj Flex. Electron.* **7**, 29 (2023).
96. Wang, Y. et al. A three-dimensional neuromorphic photosensor array for nonvolatile in-sensor computing. *Nano Lett.* **23**, 4524–4532 (2023).
97. Fu, X. et al. Graphene/MoS<sub>2-x</sub>O<sub>x</sub>/graphene photomemristor with tunable non-volatile sensitivities for neuromorphic vision processing. *Light. Sci. Appl.* **12**, 39 (2023).
98. Lee, S., Peng, R., Wu, C. & Li, M. Programmable black phosphorus image sensor for broadband optoelectronic edge computing. *Nat. Commun.* **13**, 1485 (2022).
99. Zhu, Y. et al. Non-volatile 2D MoS<sub>2</sub>/black phosphorus heterojunction photodiodes in the near-to mid-infrared region. *Nat. Commun.* **15**, 6015 (2024).
100. Tan, H. & van Dijken, S. Dynamic machine vision with retinomorphic photomemristor-reservoir computing. *Nat. Commun.* **14**, 2169 (2023).
101. Lao, J. et al. Ultralow-power machine vision with self-powered sensor reservoir. *Adv. Sci.* **9**, 2106092 (2022).
102. Sun, Y. et al. In-sensor reservoir computing based on optoelectronic synapse. *Adv. Intell. Syst.* **5**, 2200196 (2023).
103. Zhang, Z. et al. In-sensor reservoir computing system for latent fingerprint recognition with deep ultraviolet photo-synapses and memristor array. *Nat. Commun.* **13**, 6590 (2022).
104. Sun, L. et al. In-sensor reservoir computing for language learning via two-dimensional memristors. *Sci. Adv.* **7**, eabg1455 (2021).
105. Chen, J. et al. Optoelectronic graded neurons for bioinspired in-sensor motion perception. *Nat. Nanotechnol.* **18**, 882–888 (2023).
106. Wu, X. et al. Ultralow-power optoelectronic synaptic transistors based on polyelectrolyte dielectrics for in-sensor reservoir computing. *Sci. Adv.* **10**, eadn4524 (2024).
107. Leng, Y. B. et al. A near-Infrared retinomorphic device with high dimensionality reservoir expression. *Adv. Mater.* **16**, 2411225 (2024).
108. Wu, X. et al. Wearable in-sensor reservoir computing using optoelectronic polymers with through-space charge-transport characteristics for multi-task learning. *Nat. Commun.* **14**, 468 (2023).
109. Zha, J. et al. A 2D heterostructure-based multifunctional floating gate memory device for multimodal reservoir computing. *Adv. Mater.* **36**, 2308502 (2024).
110. 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).
111. Kumar, M., Abbas, S. & Kim, J. All-oxide-based highly transparent photonic synapse for neuromorphic computing. *ACS Appl. Mater. Interfaces* **10**, 34370–34376 (2018).
112. Li, G. et al. Photo-induced non-volatile VO<sub>2</sub> phase transition for neuromorphic ultraviolet sensors. *Nat. Commun.* **13**, 1729 (2022).
113. Hassan, M. Y. & Ang, D. S. On-demand visible-light sensing with optical memory capabilities based on an electrical-breakdown-triggered negative photoconductivity effect in the ubiquitous transparent Hafnia. *ACS Appl. Mater. Interfaces* **11**, 42339–42348 (2019).
114. Jang, H. et al. In-sensor optoelectronic computing using electrostatically doped silicon. *Nat. Electron.* **5**, 519–525 (2022).
115. Wu, G. et al. Ferroelectric-defined reconfigurable homojunctions for in-memory sensing and computing. *Nat. Mater.* **22**, 1430–1430 (2023).
116. Xiang, D. & Liu, T. Extending in-sensor computing from static images to dynamic motions. *Nat. Electron.* **6**, 801–802 (2023).

## Acknowledgements

This work was supported by the STI 2030-Major Projects 2022ZD0210200, National Natural Science Foundation of China under Grant 92264201, 62025111 and 62274007, and Shanghai Municipal Science and Technology Major Project.

## Author contributions

H.W. and J.T. supervised and led the review project. X.W. and H.H. wrote the main manuscript text and conducted the literature review. R.H. and Y.D. contributed to figure preparation and literature research. Y.W., B.G., and H.Q. contributed to review design and project management. All the authors prepared figures and reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Jianshi Tang or Huaqiang Wu.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025