



Cite this: *Nanoscale Horiz.*, 2025,
10, 205

Received 14th August 2024,
Accepted 11th November 2024
DOI: 10.1039/d4nh00405a
rsc.li/nanoscale-horizons

Emerging 2D materials hardware for in-sensor computing

Yufei Shi,^{†,a} Ngoc Thanh Duong^{†,ab} and Kah-Wee Ang  ^{*,a}

The advent of the novel in-sensor/near-sensor computing paradigm significantly eliminates the need for frequent data transfer between sensory terminals and processing units by integrating sensing and computing functions into a single device. This approach surpasses the traditional configuration of separate sensing and processing units, thereby greatly simplifying system complexity. Two-dimensional materials (2DMs) show immense promise for implementing in-sensor computing systems owing to their exceptional material properties and the flexibility they offer in designing innovative device architectures with heterostructures. This review highlights recent progress and advancements in 2DM-based in-sensor computing research, summarizing the unique physical mechanisms that can be leveraged in 2DM-based devices to achieve sensory responses and the essential biomimetic synaptic characteristics for computing functions. Additionally, the potential applications of 2DM-based in-sensor computing systems are discussed and categorized. This review concludes with a perspective on future development directions for 2DM-based in-sensor computing.

1. Introduction

In the era of big data and the internet-of-things (IoT), the number of sensory nodes is expected to exceed 125 billion within the next decade.¹ This scenario leads to the generation of vast amounts of unstructured and duplicated data generated by the sensory units, posing a significant burden on central processing hardware.² Consequently, the conventional von Neumann architecture, which relies on data transfer between

^a Department of Electrical & Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583, Singapore.

E-mail: eleakw@nus.edu.sg

^b Department of Energy Science, Sungkyunkwan University, Suwon 16419, Republic of Korea

† Y. S. and N. T. D. contributed equally to this work.



Yufei Shi

Yufei Shi received her BEng degree in Integrated Circuits and Integrated Systems from Xidian University, China, in 2021, and her MS degree in Electrical Engineering from the National University of Singapore in 2022. She is currently pursuing a PhD degree in Electrical Engineering at the Department of Electrical and Computer Engineering, National University of Singapore. Her research focuses on neuromorphic hardware utilizing ferroelectrics and two-dimensional (2D) materials.



Ngoc Thanh Duong

Ngoc Thanh Duong earned his Bachelor of Science and PhD from the University of Science at Vietnam National University in Hanoi in 2010 and Sungkyunkwan University in Suwon, Korea, in 2020, respectively. From 2020 to 2021, he served as a research assistant professor at Dongguk University in Seoul, Korea. In September 2021, he moved to National University of Singapore as a Research Fellow until August 2023. He was awarded the Korean BrainPool Fellowship in September 2023 and is currently a Research Professor at Sungkyunkwan University. His research interests focus on retinomorphic and neuromorphic hardware utilizing 2D ferroelectric materials.

physically separated hardware, is proving inadequate for information processing due to its substantial energy and time consumption, data storage demands, and security issues. To address these challenges, the in-sensor computing technique has been proposed, which combines sensing and computing functions. This architecture allows information to be processed directly within the multiple connected sensors, eliminating the need for data transfer and the sensor/processor interface.³

Recent advances in light-sensing technology for optoelectronics sensory devices, including phototransistors, photonic memory, photovoltaic, gas sensing, and piezoelectric-based tactile sensor devices, have become a major focus in in-sensor device research. State-of-the-art devices based on 2D nanostructured materials for various applications are summarized in Fig. 1. In-sensor computing, which integrates computational capabilities directly within the sensor, enables local data processing rather than relying on an external processor. This approach can significantly enhance data processing efficiency and speed, reduce latency, and lower power consumption.

1.1 Different categories of in-sensor computing models and progress

This review categorizes in-sensor computing into four primary domains: retinomorphic hardware, circuit integration and materials, light detection and ranging (LiDAR), and multimodal sensing and processing. Advances in light-sensing technology have led to the development of retina-inspired image sensors for artificial intelligence (AI) photonic chips.¹² Current retinomorphic hardware, based on electrostatically doped photodiodes, involves numerous interconnections within the circuit.^{13,14} Additionally, charge-coupled devices (CCDs) face challenges in physical implementation due to the need for stacking multiple van der Waals layers of ambipolar materials to perform photon capturing and carrier memorizing functions.^{11,15} Emulating the structure and function of retinal cells is vital for developing advanced visual processing capabilities in sensors.^{5,16–18} Moreover, visual

adaptation techniques that enable sensors to adjust to varying visual environments, akin to the adaptability of biological retinas, are essential for enhancing sensor performance in diverse conditions. To emulate the function of biological rod and cone cells in photoreceptors, oxide-induced trap MoS₂ photodiodes⁴ and perovskite-based homogeneous integration^{19,20} are employed to enable effective adaptation when processing information under dim or bright light.

Significant focus is directed toward utilizing two-dimensional (2D) semiconductor materials to develop sensor circuits that are both more efficient and compact. Among these materials are two-dimensional WSe₂, PtSe₂, MoTe₂/PdSe₂ heterostructures, and electrostatically doped silicon, each capitalizing on their inherent ambipolar transport properties.^{21–23} Additionally, integrating ferroelectric materials, known for their non-volatility and low power consumption, is paramount for advancing in-sensor computing technologies. The literature explores the ferroelectricity and photoconductivity of two-dimensional materials, such as α -In₂Se₃ and SnS₂, to develop innovative retinomorphic sensors.^{5,24–26} These ferro-photonics AI chips leverage switchable polarization dipoles, enabling electrical programming while optical stimulation is applied to the channel to modulate the polarization state, thereby facilitating multifunctional operation.

The fundamental mechanism of a typical LiDAR sensor involves transmitting laser signals and analyzing the reflected signals to determine the distance to an obstacle or collision.²⁷ Optical stimulation can be transmitted in either two forms: as pulses or as frequency-modulated continuous waves (FMCW). The study of flying animals, such as bats, which use natural sonar, informs the development of LiDAR (light detection and ranging) systems for enhanced spatial awareness and navigation.⁹ Inspired by how bats navigate using echolocation, LiDAR is used to penetrate dense forest canopies and map the underlying terrain, uncovering hidden archaeological sites beneath dense vegetation. This application is crucial for forestry management, environmental monitoring, and habitat characterization. Additionally, some motion detection models are inspired by flying insects, such as *Drosophila* (fruit flies) and locusts, which use lobula giant movement detector (LGMD) neurons in their eyes to detect motion within milliseconds, serving as a model for designing motion detection algorithms in LiDAR systems.^{8,28}

Multimodal processing in in-sensor computing refers to the integration and simultaneous processing of multiple types of sensory data within a single sensor system. This approach enables the sensor to capture and analyze different forms of information, such as visual, auditory, and tactile. For instance, a multimodal in-sensor computing system might combine image data from a camera sensor with depth information from a LiDAR sensor and sound data from a microphone. Sensors that can process tactile and taste information expand the range of detectable sensory inputs, provide comprehensive analysis in real-time, reduce data transfer to external processors, and enhance overall system efficiency. Integrating these data types allows the system to understand the environment more



Kah-Wee Ang

Kah-Wee Ang is a tenured Associate Professor of Electrical and Computer Engineering at the National University of Singapore (NUS). Concurrently, he is a Chief Technology Officer at the National Semiconductor Translation and Innovation Centre (NSTIC) and serves as Deputy Director of E6 NanoFab, a micro- and nanofabrication research center at NUS. His research centers on advancing nanoelectronics and photonics technologies, with a particular focus on developing neuromorphic computing systems based on two-dimensional (2D) materials for artificial intelligence applications.

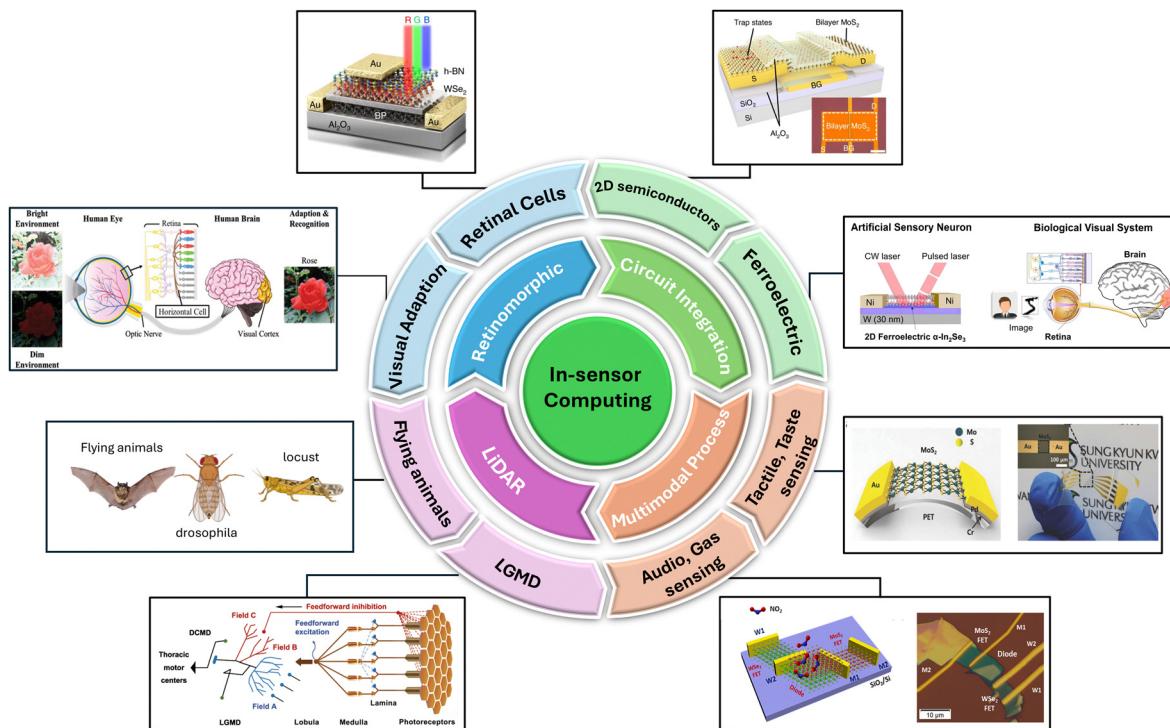


Fig. 1 Overall graphic review of progress and perspectives for in-sensor computing, including four different mainframes: retinomorphic hardware, circuit integration & materials, light detection and ranging (LiDAR), and multimodal processing. Reproduced with permission.⁴ Copyright 2024, Springer Nature Limited. Reproduced under the terms of Creative Commons CC BY License.⁵ Copyright 2023, The Authors. Advanced Science published by Wiley-VCH GmbH. Reproduced with permission.⁶ Copyright 2018, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. Reproduced with permission.⁷ Copyright 2021, American Chemical Society. Reproduced under the terms of Creative Commons CC BY License.⁸ Copyright 2021, The Author(s), published by Springer Nature. Reproduced under the terms of Creative Commons CC BY License.⁹ Copyright 2021, The Authors, published by Frontiers. Reproduced with permission.¹⁰ Copyright 2023, Wiley-VCH GmbH. Reproduced with permission.¹¹ Copyright 2021, The Author(s), under exclusive license to Springer Nature Limited.

accurately. This capability is advantageous in applications such as autonomous vehicles, where combining visual and depth information can improve object detection and navigation, or in smart devices that need to interact with users more intuitively and responsively.

1.2 The essence of in-sensor computing for practical data science and machine learning

We encapsulate the multifaceted nature of in-sensor computing, emphasizing its interdisciplinary approach by integrating concepts from biology, materials science, and electronics, as highlighted in Fig. 2. This unique aspect of in-sensor computing showcases how innovations in these areas contribute to the development and enhancement of in-sensor computing systems. In-sensor computing technology represents a significant advancement in image processing and computer vision. As depicted in the figure, this technology integrates computation directly within the image sensor, contrasting sharply with the conventional sensory approach, where raw image data is first transferred to a central processing unit (CPU) or a graphics processing unit (GPU) for subsequent processing. The traditional image processing pipeline typically involves several stages: image capture by the sensor, analog-to-digital (A/D) conversion, temporary storage in memory, and data transfer to a CPU or GPU for

further processing and classification.^{29–31} While effective for commercialization, the continuous shuttling of data between different components incurs substantial latency and power consumption. In contrast, in-sensor computing eliminates these inefficiencies by enabling direct data processing within the sensor itself. This approach reduces latency and power consumption, streamlines the image processing workflow, and offers significant improvements in efficiency and performance over traditional methods.

The essence of in-sensor computing lies in embedding certain levels of processing directly within the image sensor, thereby enabling early-stage processing and computing at the point of capture or at the edge where data is generated. The image sensor, which consists of an array of photodetectors, performs initial computations on the raw image data as it is captured. These computations can be categorized into low-level and high-level processing tasks. Low-level processing involves fundamental image processing operations, such as edge detection, noise reduction, and filtering, which are computationally intensive and require significant data movement in conventional systems. By integrating these functions into the sensor, the amount of data that needs to be transferred to subsequent hardware is significantly reduced. For instance, edge detection can be performed within the sensor array, producing a

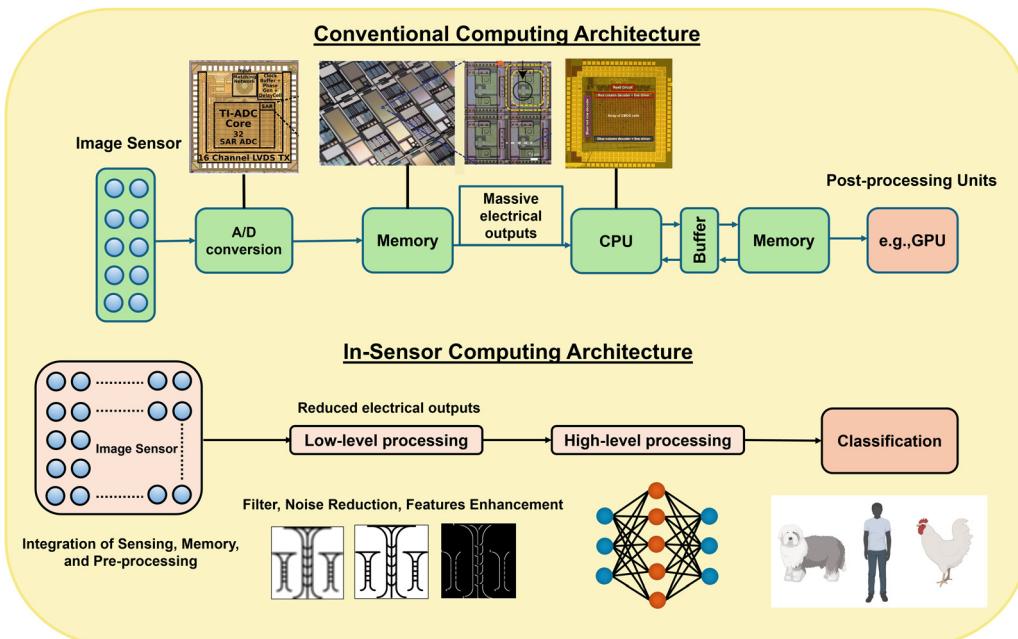


Fig. 2 Block diagrams depict the comparison between conventional imaging sensor architecture and in-sensor computing technology. Reproduced under the terms of Creative Commons CC BY License.²⁹ Copyright 2017, The Author(s), published by Springer Nature. Reproduced under the terms of Creative Commons CC BY License.³⁰ Copyright 2018, The Author(s), published by Springer Nature. Reproduced with permission.³¹ Copyright 2021 John Wiley & Sons, Ltd.

processed output that highlights contours and edges within the scene, as illustrated by the progression from the raw image to the processed edge map in Fig. 2. This approach economically reduces data bandwidth and accelerates processing time.

Building on low-level processing, high-level processing tasks, including sophisticated functions like pattern recognition and motion detection, are executed. High-level processing aims to identify and categorize objects within the image, leveraging machine learning algorithms and neural networks. In the context of in-sensor computing, these algorithms can be partially or fully implemented within the sensor hardware, allowing for rapid real-time identification and classification of objects. For example, the processed data from low-level edge detection can be used to extract specific features, such as shapes and textures, which are then used to classify objects into different categories. The final output of these processing stages is a labeled image indicating the presence and type of objects detected within the scene.

In-sensor computing technology leverages advanced fabrication techniques and sensor architectures to embed computational capabilities within the pixel array. This involves using specialized hardware such as analogue processing circuits, digital signal processors, and emerging devices like memristors and photonic processors, which can perform computation at the speed of light. These innovations enable the sensor to handle a substantial computation, offloading work from the central processor and reducing overall system power consumption.

Moreover, in-sensor computing supports parallel processing due to the inherent parallelism of the sensor array. Each pixel can operate independently, performing computations concurrently

with its neighbors, significantly accelerating the processing speed. This parallelism contrasts sharply with the sequential processing nature of conventional CPUs, making in-sensor computing highly efficient for image processing tasks.

1.3 Comparative overview of human nervous systems and artificial sensory systems

The connection between human nervous and artificial sensory systems is based on their basic operational principles. Both rely on specialized receptors to detect environmental stimuli and convert them into interpretable signals, such as electrical or chemical spikes for further processing units (Fig. 3). In humans, sensory receptors such as photoreceptors and mechanoreceptors transform various physical or chemical stimuli into electrical impulses, which are then transmitted to and processed by the brain. Similarly, image and tactile sensors in artificial systems convert similar stimuli into digital signals processed by a central processing unit (CPU).

The retina has a laminar organization, comprising Ganglion, inner nuclear, photoreceptor, and pigmentary layers.³² Light-sensitive cells (photopigments) in the disk membranes absorb light, triggering changes in the photoreceptor membrane potential.³³ This change causes the synaptic terminal in the photoreceptor to encode the information in a spiking format, which is then transferred to the brain's visual cortex via the optic nerve for processing. The retina contains two types of photoreceptors, distinguished by the morphology of their outer segments. Rod photoreceptors have long, cylindrical outer segments containing numerous disks, while cone photoreceptors have shorter, tapering outer segments with fewer disks. Rods are over 1000 times more sensitive to light than cones.

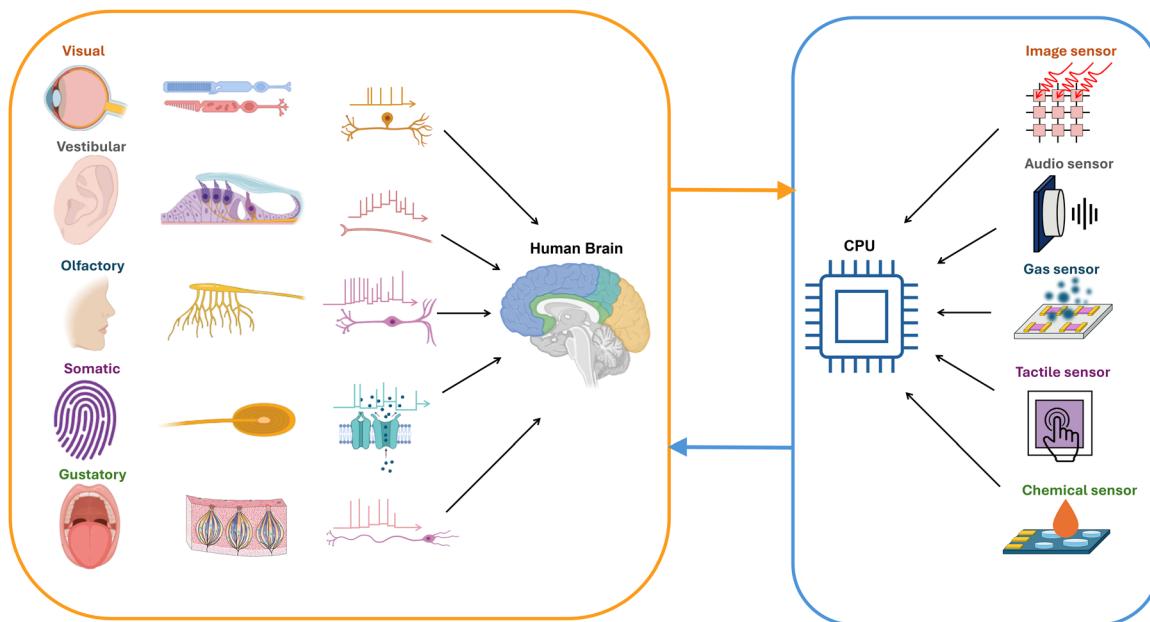


Fig. 3 A comparative overview of human nervous systems and their artificial counterparts, illustrating the pathways from sensory receptors to the central processing units in both biological and artificial systems.

due to their increased number of disks and higher photopigment concentration. Each human retina is estimated to contain approximately 5 million cones and 92 million rods.³⁴ This mechanism is analogous to an artificial image sensor, which converts light into digital signals and encodes them into electrical spikes or synaptic excitatory/inhibitory signals for further processing by the CPU.

A similar approach is found in auditory and vestibular systems. In the auditory systems, sound waves, understood as the motion of air pressure, move the tympanic membranes. This movement passes through the ossicles, oval windows, and cochlea fluid, ultimately causing a response in sensory neuron hair cells. The human auditory system can respond to sound wave frequencies ranging from 20 Hz to 20 000 Hz.³⁵ Analogous to hair cells in the inner ear, audio sensors capture sound wave information and transmit it to a processor for further analysis.

The chemical sense ability in humans, encompassing gustation (taste) and olfaction (smell), detects environmental chemicals. Smell and taste are directly connected to fundamental internal needs such as thirst, hunger, emotion, sexual behavior, and specific forms of memory.³⁶ The chemically sensitive part of a taste receptor cell is its apical end, located near the tongue's surface. These apical ends have microvilli that extend into the taste pore, exposing the cell to mouth contents. Although taste receptor cells are not neurons by standard criteria, they form synapses with gustatory afferent axons and basal cells, creating a simple information-processing circuit within each taste bud. When a chemical activates a taste receptor cell, its membrane potential changes, typically depolarizing, leading to the opening of voltage-gated calcium channels and transmitter release.³² Sour and salty taste cells release serotonin, while sweet, bitter, and umami cells release adenosine

triphosphate (ATP), which excites the postsynaptic sensory axon and causes action potentials that communicate the taste signal to the brain stem. Taste cells may also use other transmitters like acetylcholine, gamma-aminobutyric acid (GABA), and glutamate, though their functions are unclear. Unlike taste receptors, odorants or chemical stimuli in the air dissolve in the mucus layer before reaching olfactory receptor cells. Olfactory transduction involves odorants binding to receptor proteins, stimulating a G-protein, activating adenylyl cyclase, and forming cyclic adenosine monophosphate (cAMP), which opens cation channels, allowing Na^+ and Ca^{2+} influx and subsequently opening Ca^{2+} -activated Cl^- channels, leading to membrane depolarization. If the receptor potential is large enough, it triggers action potentials that propagate to the central nervous system. The olfactory response can terminate due to odorant diffusion, enzymatic breakdown, or receptor cell adaptation, which reduces the response over time, even in the presence of the stimulus.

Mechanoreceptors, mainly present in somatic sensory systems, respond to physical distortion such as stretch, pressure, or bending. The *Pacinian corpuscle* is the most extensively studied mechanoreceptor, located deep within the dermis, with dimensions reaching 2 mm in length and nearly 1 mm in diameter. It converts physical stimuli into electrical signals that the nervous system can interpret. The functioning of the human somatic sensory system is comparable to artificial tactile sensors, which predominantly utilize piezoelectric materials such as lead zirconate titanate (PZT),³⁷ barium titanate (BaTiO_3),³⁸ and zinc oxide (ZnO) to convert mechanical force into electrical signals.³⁹

Both biological and artificial systems exemplify a modular approach to sensory processing, where distinct types of receptors or sensors are specialized for specific stimuli, ensuring

efficient and accurate environmental perception and response. The CPU in artificial systems parallels the integrative and interpretative functions of the human brain, highlighting the interdisciplinary convergence of biology, neuroscience, and engineering in the development of sensory technologies.

In this review, we thoroughly examine recent advancements in in-sensor computing hardware utilizing emerging two-dimensional materials (2DMs). We categorize and consolidate the unique properties and physical mechanisms of 2DMs that enable in-sensor computing. To provide a comprehensive understanding of the in-sensor computing paradigm, we conclude its evolution process and outline the essential characteristics required for high-performance hardware. In addition, various applications possibly realized by 2DM-based in-sensor computing hardware are categorized, such as in-sensor reservoir computing, motion detection, and multimodal signal processing. Finally, a summary with current research efforts, key challenges, and potential directions for future exploration in this field is provided.

2. Possible mechanisms leveraged in 2DMs in-sensor computing hardware

Emerging two-dimensional materials exhibit superior properties across various aspects, paving the way for innovation in in-sensor computing technology. Firstly, the thickness of 2DMs can be scaled to a few or even a single atomic layer,^{40,41} endowing them with exceptional electrical, mechanical, and optical characteristics not observed in their bulk counterparts.⁴² This atomically thin dimension results in a high surface-to-volume ratio, greatly enhancing their interaction with external stimuli and leading to high sensitivity and rapid response times. Integrating 2DMs into devices significantly shrinks device dimensions to the sub-nanometer scale, facilitating high-density integration and enabling the implementation of more powerful functions with smaller circuit-level footprints. For semiconductors 2DMs, notorious issues such as gate leakage current and the short-channel effect can be suppressed when integrated into transistor configurations due to the excellent quantum confinement of carriers in atomically thin channels and uniform gate modulation of channel carriers.^{43–45}

Secondly, the electronic properties of 2DMs can be precisely tuned through various methods such as defect engineering, chemical doping, and strain application.^{4,6} This tunability enables the optimization of 2DMs for specific in-sensor computing functions. For example, leveraging the tunable bandgap of 2DMs can lead to broadband vision systems with robust photoresponse to near-infrared (NIR) light, which is crucial for applications like night vision and medical imaging.^{15,46–50} Reconfigurable, van der Waals infrared photodetectors can perform *in situ* edge extraction at the detection end, reducing the transmission burden by extracting key features directly within the sensor and enhancing the efficiency of image perception systems.⁵¹

Furthermore, the absence of dangling bonds in 2DMs leads to ideal periodic structures and uniform charge distribution,

enabling the formation of clean interfaces in 2DM-based devices. By leveraging the weak interlayer interactions within 2DMs, various heterostructures can be constructed by stacking multiple 2DMs.^{52–55} This approach offers opportunities for designing novel device architectures that realize multifunctionality or precise control of specific functions,⁵⁶ enabling devices to perform complex computing tasks directly at the sensor interface. Additionally, heterostructures of different 2DMs can exhibit synergistic effects that enhance their overall performance.

The rapid development of 2DM synthesis technology, such as chemical vapor deposition, enables the production of large-scale, high-quality 2DMs suitable for industrial applications.^{57–60} This compatibility facilitates the seamless incorporation of 2DMs into current manufacturing workflows, accelerating the development and deployment of in-sensor computing technologies. Moreover, their excellent mechanical flexibility and robustness allow integration into various substrates, further manifesting their potential in flexible and wearable electronics. Beyond these shared advantages, different 2DMs also possess unique properties and exhibit exceptional responses to specific external sensory signals or combinations of stimuli, demonstrating significant potential for in-sensor computing.

2.1 Ambipolar photovoltaic effect

Optical information is a ubiquitous sensory source in our daily lives. Sensing and processing light stimuli across varying wavelengths is a primary branch of in-sensor computing systems, with the photovoltaic effect being the most widely adopted mechanism in 2DM-based in-sensor computing systems. The photovoltaic effect generates photocurrent inside the device upon exposure to light; the process involves creating electron-hole pairs when photons strike the material. These photo-generated carriers are separated by the internal electric field, producing a photocurrent. Due to their reduced dimensionality, various monolayer transition metal dichalcogenides (TMDs) (e.g., MoS₂ and WS₂) have direct bandgaps, facilitating efficient light absorption and enhanced light interaction, making them highly suitable for optical information processing scenarios.⁶¹ This strong light absorption leads to high photosensitivity, making 2DM-based devices ideal for detecting low-intensity light signals and enabling them to function in dim environments and sense subtle changes in illumination. In addition, the high carrier mobility of 2DMs allows rapid transport of photogenerated carriers, minimizing recombination losses and enhancing the overall efficiency of photovoltaic devices. This characteristic translates into fast response time during device operation, essential for real-time sensing and computing applications. Moreover, the gate-tunability of photocurrent provides another modulation terminal for optical signal processing. Applying an external electric field allows precise control over the photoresponse characteristics, enhancing the device's functionality in dynamic computing applications and expanding the spectrum's detecting range. Furthermore, heterostructure configurations among these photosensitive 2DMs further boost the capability of 2DM-based optical information processing.^{22,62,63} The heterostructures formed by

stacking different 2DMs create various tailored electrical and optical properties and can exhibit unique photovoltaic effects due to the synergistic interactions between the layers, promoting efficient charge separation and collection.^{64–67}

For instance, the ambipolar open circuit voltage (V_{oc}) observed in the gate-tunable junction mode of broken-gap SnSe₂/MoTe₂ heterostructure exemplifies this potential.⁶⁸ Harnessing the bi-direction photovoltaic effect of van der Waals (VdW) heterostructures, the 2DM-based in-sensor computing system exhibits superiority in broadband convolutional processing, outperforming conventional CMOS technology for broadband sensing and convolutional processing.^{22,69} Pi *et al.* demonstrated a practical broadband convolution processing approach using the gate-tunable linear broadband photovoltaic effect of PdSe₂/MoTe₂ heterostructures (Fig. 4a).²² Under different gate voltages, the band alignment between n-type PdSe₂ and p-type MoTe₂ varies, resulting in distinct positive and negative photovoltaic responses. Negative gate voltages induce type-II band alignment, promoting the separation of photogenerated electron-hole pairs through a built-in electric field. In contrast, positive gate voltages create a type-III band alignment, introducing a tunneling barrier that modulates the current flow (Fig. 4b). This configuration presents a linear relationship between photoresponse and both light intensity and gate voltage (Fig. 4c), ideal for implementing BCP and enabling the direct integration of convolution operations within the sensor, facilitating efficient and low-power image processing. The device shows efficient sharpness operation and edge enhancement

for Jasper Ridge datasets images containing information across different wavelength bands, achieving a significant improvement of over 90% in image recognition tasks compared to single band-based convolutional neural networks.

In addition to conventional static optoelectronic convolutional processing, Yang *et al.* employed the graphene/Ge heterostructure to construct an in-sensor dynamic computing system for robust tracking of dim targets (Fig. 4d).⁶³ By dynamically tuning the photoresponse *via* adjustable gate voltages, a dynamic kernel is implemented using an optoelectronic device array that contains internal correlations and can respond to spatial light patterns. As shown in Fig. 4e, the array comprises both passive and active components, where passive devices are sensitive to local light intensity, and active devices' responses are modulated by both local and surrounding light intensities, facilitating the selective amplification of minor light intensity variations. This configuration allows for robust extraction of edge features from low-contrast images, a critical capability in complex visual environments. Effective edge extraction from low-contrast scenes (Fig. 4f) and improved recognition accuracy of human faces across varying contrast levels are achieved with this intelligent machine vision system.

2.2 Ferroelectricity

2.2.1 Integration of 2D semiconductor and ferroelectric dielectrics. Conventional dielectric ferroelectric materials have received significant attention due to their fast dipole switching, nonvolatile polarization hysteresis, and tunable polarization

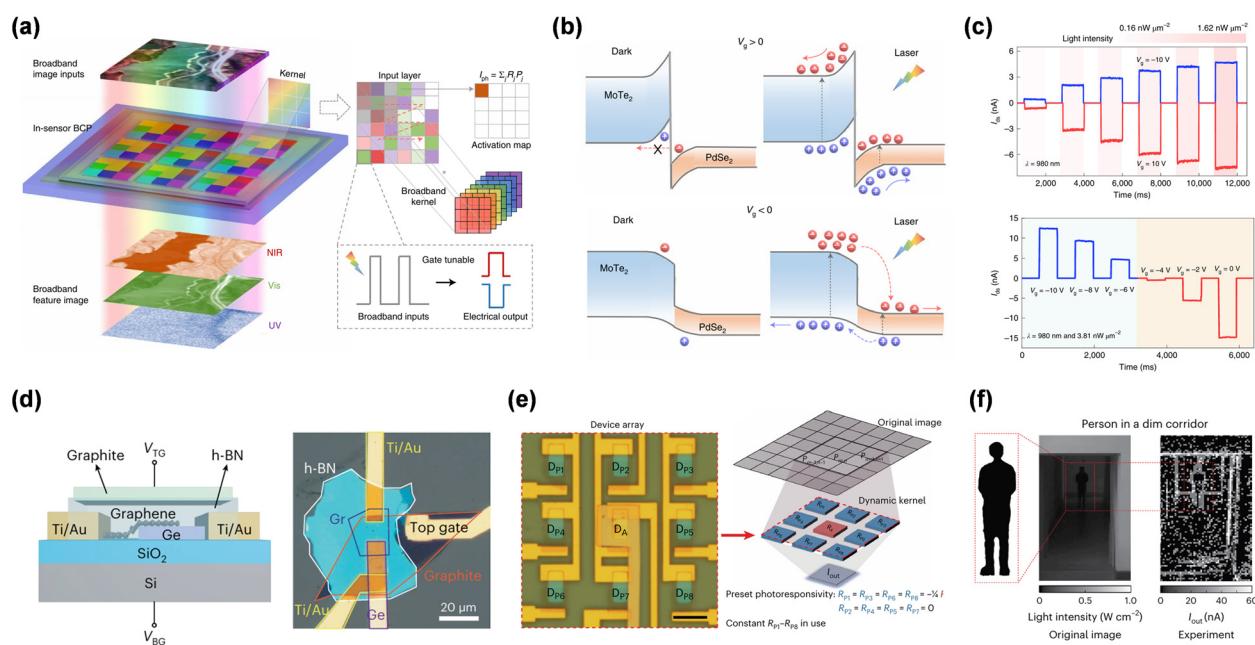


Fig. 4 Photovoltaic effect adopted in in-sensor computing systems. (a) Schematic of broadband convolutional processing (BCP), dynamic kernels inside BCP are composed of PdSe₂/MoTe₂ heterostructure with gate-tunable photoresponsivity. (b) Different electron–hole pair generation patterns under light illumination, arising from different band-alignment caused by opposite gate voltages. (c) Light-tunable and gate-tunable current response. Reproduced with permission.²² Copyright 2022, Springer Nature Limited. (d) Schematic of graphene–Ge heterostructure device. (e) Edge features extraction of low-contrast images by using graphene–Ge heterostructure array. (f) Effective extraction edge profile of dim image using in-sensor dynamic computing. Reproduced with permission.⁶³ Copyright 2024, Springer Nature Limited.

states, demonstrating significant potential for multifunctional hardware implementations.⁷⁰ Oxide-based ferroelectric field-effect transistors (*e.g.*, HZO, PZT, and BaTiO₃) have shown memory capabilities and are emerging as promising candidates for overcoming the limitations of traditional von Neumann computing, including low processing speed and high energy consumption. The integration of 2D materials with these dielectric ferroelectric materials further enhances the device functionality.^{71–75} This combination paves the way for their application in in-sensor computing due to the unique properties of 2DMs, such as high photosensitivity, which can be effectively associated with ferroelectrics.

For instance, by exploiting the ferroelectricity in hafnium-zirconium oxide (HZO) and semiconducting two-dimensional p-type black phosphorus (BP), Hao *et al.* demonstrated a 2D optoelectronic memristive device with superior electrical properties and synaptic functions.⁷⁶ This strategic implementation leverages the ferroelectric properties of HZO to modulate the energy band structure and electron transport in the BP channels, resulting in enhanced photodetection and memristive behavior. The broadband optical response of black phosphorus and the tunable conductance *via* HZO regulation underscore the device's potential for highly efficient, multifunctional in-sensor computing applications.

Furthermore, conventional hafnium-zirconium oxide not only possesses nonvolatile ferroelectricity, but with the careful engineering of material composition (*i.e.*, increasing the proportion of Zr in HZO), volatile anti-ferroelectricity can also be achieved. This enhances the capability of oxide-2DM integrated devices for handling temporal information. By exploiting the inherent fading memory (short-term effect) in Hf_{0.25}Zr_{0.75}O₂, Shi *et al.* demonstrated a multi-modal reservoir computing (RC) system for efficient spatiotemporal information processing.⁷⁷ The dipoles polarize under the control of an external electric field and depolarize spontaneously upon its removal. This dynamic switching process regulates carrier flow inside the MoS₂ channel and adjusts the output current, enabling the generation of diverse reservoir states with different combinations of pulse sequences. Besides the electrical modulation, the RC system also harnessed the light sensitivity induced by the 2D MoS₂ integration to achieve control of temporal dynamics under optical stimuli. This dual-mode modulation response significantly enhances the reservoir state richness and boosts the overall system performance, resulting in an impressive accuracy of 95.4% in spoken digit recognition tasks.

2.2.2 Two-dimensional ferroelectrics-based FeFET. Building on the previous discussion, the combination of 2DM with conventional ferroelectric dielectric materials enhances the performance of 2D devices. However, this configuration has its limitations. For instance, ferroelectric dielectrics have a critical thickness of around 10 nm, which constrains the high-density integration and miniaturization of ferroelectric field-effect transistors (FeFETs).⁷⁸ When the ferroelectric layer is scaled below this thickness, a disordered polycrystalline microstructure appears, leading to reduced long-range Coulomb coupling and an increased depolarization field.^{79,80} Additionally,

integrating 2DM with ferroelectric dielectrics requires the epitaxial growth of ferroelectric thin films. The lattice mismatch between the two layers often generates unfavorable defect states and causes surface reconstruction.^{81–83} Furthermore, the large depolarization field induced by incomplete charge compensation in the ferroelectric layer raises concerns about device retention and endurance.^{84,85}

In contrast, the 2D ferroelectrics exhibit a small critical thickness and enhanced immunity to functional fatigue, facilitating ultra-fast polarization flipping, a flat surface, and controllable ferroelectric polarization. Beyond serving as dielectric layers, 2D ferroelectric materials display diverse properties, ranging from insulators (*e.g.*, CuInP₂S₆) to semiconductors (*e.g.*, α -In₂Se₃) and metals (*e.g.*, WTe₂), significantly expanding their application scenarios.⁷⁸ Moreover, forming a clean interface with heterogeneous integration between 2D ferroelectrics and other 2D layers allows for active modulation of the optical and electrical properties of devices.

CuInP₂S₆ (CIPS) is widely utilized as the dielectric layer in fully two-dimensional ferroelectric field-effect transistors due to its robust switchable ferroelectric polarization at room temperature. The ferroelectricity properties arise from the displacement of Cu ions within its layered structure, where these ions occupy off-center positions within the octahedral voids formed by the surrounding P and S atoms.^{86,87} This displacement creates an electric dipole, leading to spontaneous polarization that can be switched by an external electric field. The ferroelectricity of CIPS is preserved even at the monolayer limit and exhibits a relatively high Curie temperature of approximately 340 K.⁸⁸ Integrating CIPS with a 2D semiconductor channel layer allows modulation of the hysteresis loop in dual-swept channel conductivity *via* ferroelectric polarization, enabling both sensing and memory functions within the same device. Additionally, CIPS's large polarization and a strong coupling between its ferroelectric and mechanical properties make it highly suitable for applications in flexible electronics.

Wang *et al.* report a 2D heterostructure-based ferroelectric field-effect transistor that uses light-modulated ferroelectric polarization to realize a multi-functional vision system.⁸⁹ When illuminated by photon energy, the interfacial charges at the SnS₂/CIPS interface are compensated by photogenerated carriers, resulting in the change of built-in field and polarization direction. This modulated polarization in CIPS causes the depletion or accumulation of electrons in the SnS₂ channel, switching the device between high resistance state (HRS) and low resistance state (LRS) (Fig. 5a). This photoelectric modulation property enables the device to simulate associative learning and light adaptation behaviors. The intensity and projection time of the light is crucial in controlling synaptic behavior, with longer light pulses resulting in longer retention times and a transition from short-term plasticity (STP) to long-term plasticity (LTP), as shown in Fig. 5b. Multi-level conductance states can be achieved with varying doses of light irradiation (Fig. 5c). Utilizing both short-term plasticity and long-term plasticity features, a comprehensive reservoir computing

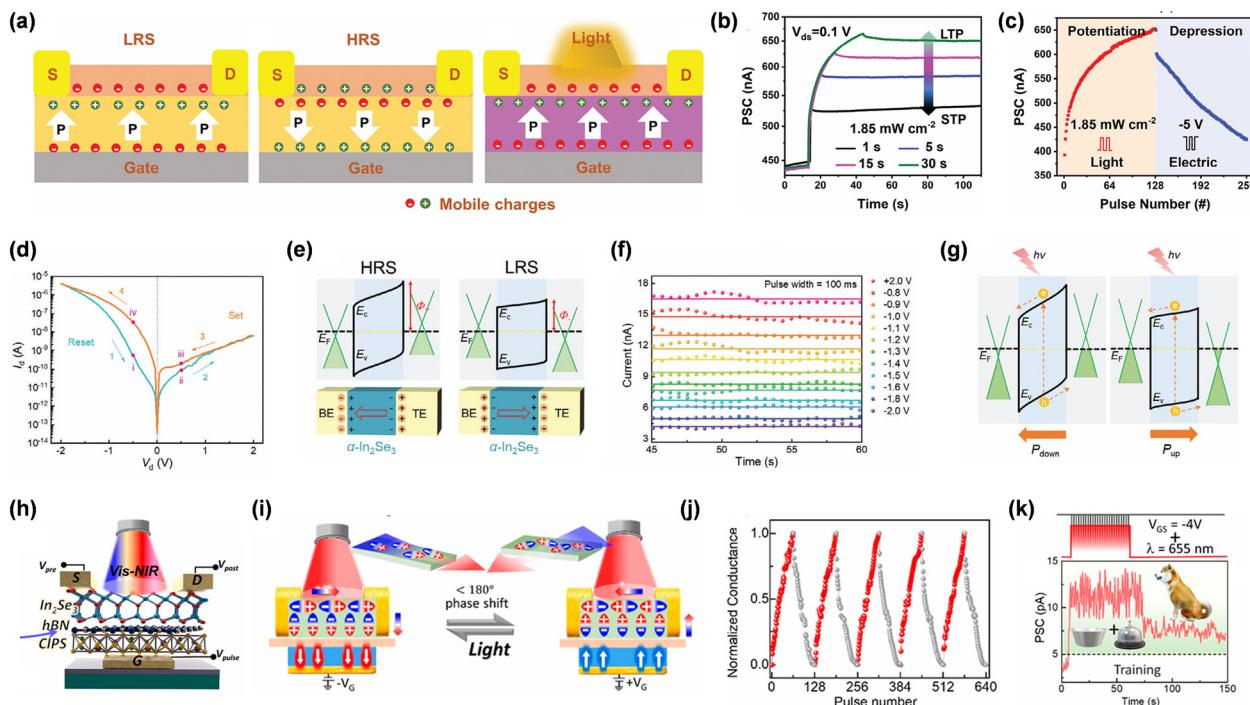


Fig. 5 2D ferroelectrics for in-sensor computing. (a) Change of channel carrier concentration by both electric and light-induced ferroelectric polarization change. (b) The transition of synaptic behavior from STP to LTP by increasing light pulse duration. (c) Multi-level PSC states are obtained by successive optical and electrical spike pulse stimuli. Reproduced under the terms of Creative Commons CC BY License.⁸⁹ Copyright 2023 The Authors. Advanced Science published by Wiley-VCH GmbH. (d) Hysteresis loop presents in output characteristics of α -In₂Se₃ device, resulting from OOP ferroelectric polarization. (e) Energy band diagrams of polarization state modulation on Schottky barrier height. Varying device states are generated with the change of polarization direction. (f) Long-term stable distinct intermediate states are achieved through electrical pulses of varied amplitudes. (g) Influence of ferroelectric polarization on the movement of photo-generated carriers under light illumination. Reproduced with permission.⁹⁰ Copyright 2024, Wiley-VCH GmbH. (h) Schematic of vdWH FeFET (CIPS/hBN/ α -In₂Se₃). (i) Light-induced dipole polarization change in α -In₂Se₃ layer. (j) Stable LTP behavior, with optical potentiation and electrical depression. (k) Successful demonstration of Pavlov's experiment implies the classical conditioning obtained from the 2D-ferroelectric-based optoelectronic synaptic device. Reproduced with permission.⁹¹ Copyright 2023, American Chemical Society.

system is physically constructed based on the demonstrated FeFET. This system includes a reservoir layer capable of capturing external temporal optical information and a fully connected layer for simultaneous training and processing. This configuration achieves a high accuracy of 93.62% in recognizing MNIST handwritten digits.

Beyond being utilized as a dielectric layer, 2D ferroelectric α -In₂Se₃ can also function as a channel layer due to its natural semiconducting properties.^{92,93} Unlike conventional FeFETs, where polarization charges accumulate at the ferroelectric and semiconductor interface, semiconductor FeFETs using α -In₂Se₃ as the channel layer exhibit charge accumulation at both the top and bottom surfaces. This approach maintains the ferroelectric property while allowing the substitution of the dielectric layer with other high-quality materials, alleviating non-ideal interface issues. The α -In₂Se₃ layer's quintuple-layer structure, with asymmetric Se atoms positions breaking centrosymmetric, easily produces spontaneous in-plane and out-of-plane polarization.⁹⁴ The covalent bond configuration causes the interlocking effect among dipoles, significantly stabilizing ferroelectric polarization even to monolayer thickness.⁹⁵⁻⁹⁷ Moreover, α -In₂Se₃'s superior photosensitivity and light-induced polarization change further enhance its capability for

integrating sensing and processing functions within a single device.

The integration of electrical and optical modulation of α -In₂Se₃'s ferroelectric polarization produces various dynamics, significantly enhancing its information processing capabilities. Leveraging the crystal regulation of α -In₂Se₃ ferroelectricity, Zeng *et al.* developed a multisensory synapse that merges sensing, computing, and memory within a single device. This system performs both in-sensor front-end processing and subsequent post-processing.⁹⁰ The dynamic modulation of retention and plasticity is linked to the adjustment of Schottky barrier height in response to ferroelectric domain motions under both light and electric stimuli (Fig. 5d). The applied electric field shifts the Fermi level of the bottom graphene layer, causing the Schottky barrier height to increase (decrease) when α -In₂Se₃ polarization orients downward (upward), resulting in the device's HRS (LRS) (Fig. 5e). Moreover, various parameters, such as pulse amplitude and width, alter the overall polarization state and direction, enabling the transition between STP and LTP and producing multiple intermediate conductance states, beneficial for synaptic characteristics (Fig. 5f). The electric field-induced Schottky barrier height change further generates a built-in field, influencing its intensity

and accelerating or decelerating the separation of photogenerated carriers under light stimuli (Fig. 5g). Consequently, the non-volatile conductance and photoresponsivity are associated with the dynamics of ferroelectric polarization within the α -In₂Se₃ layer, enhancing the multifunctionality of neural networks.

In addition, in Fig. 5h, Das *et al.* demonstrated a ferroelectric field-effect transistor composed of CIPS and α -In₂Se₃ van der Waals heterostructures, capable of sensing, storage, and processing of optical signals within a single device.⁹¹ The spontaneous dipole polarization is bidirectionally locked and controlled within the α -In₂Se₃ channel by gate voltages, enabling robust memory storage. Light-induced polarization within α -In₂Se₃ alters the dipole order in response to optical signals, facilitating optical sensing and the transition from short-term to long-term memory (Fig. 5i and j). The coupled in-plane (IP) and out-of-plane (OOP) polarizations, modulated by electrical and optical inputs, underpin complex functionalities and enable the device to mimic retinal functions, combining sensing, processing, and memory within a single unit. The system achieves high paired-pulse facilitation, color recognition, and accurate pattern recognition (92.5% with MNIST datasets), along with demonstrating optical logic operations and event learning tasks, such as Pavlov's dog experiment as shown in Fig. 5k.

2.3 Ionic filament-assisted photoconductivity

In in-sensor computing systems built using two-dimensional materials-based two-terminal memristors, the formation of conductive filaments is the most prevalent mechanism for achieving effective conductance state modulation. These filaments are typically formed through the migration of ions generated by active metal electrodes or the drift and diffusion of intrinsic vacancies within the materials under an applied electric field. Prior to the application of the external electric field, devices show high resistance due to the insulating or semiconducting nature of 2DMs between the two electrodes. Active metal atoms are oxidized at one electrode and migrated towards the other side, bridging two electrodes with conducting filaments that allow metal atoms to penetrate. These formed localized conductive paths provide a low-resistance route for current flow and alter the material's electrical properties, altering the material's electrical properties and transitioning the device to a low-resistance state. More importantly, opposite external electric fields can erase these ionic conductive paths to recover high resistance states.

Furthermore, the source of migrated ions does not only originate from metal electrodes. Several studies concluded that 2DMs inherently contain intrinsic defects or vacancies, which can also facilitate the migration of native ions (*e.g.*, sulphur vacancies in MoS₂, selenide vacancies in PdSe₂). The formation and rupture (deformation) of conductive filaments induce nonvolatile or volatile (for spontaneously ruptured filaments) characteristics into the devices, thereby introducing various synaptic plasticity.

Besides intrinsically forming conductive filaments in the 2DM itself, 2DMs also exhibit effective regulation of conductive

filaments when integrating them with other oxide layers. Wang *et al.* employ the high hydrophilicity of the 2D MXene layer and further integrate it with the photo-active ZnO layer to construct a multi-modal in-sensor computing system that emulates the capability of human perception system for sensing and processing visual information in a complex environment.⁹⁸ Inside the MXene-ZnO memristor, oxygen vacancies migrate and form a conductive filament under the modulation of applied voltage; simultaneously, due to the existence of abundant hydrophilic sites in the MXene nanosheet (Fig. 6a), the device shows high protonic sensitivity and the electrostatic attraction between proton and oxygen vacancies causes restriction of the conductive filament growth (Fig. 6b). Therefore, the electric field and relative humidity control and modulate the resistive switching process of the memristor and enable an adaptive behavior sensitive to both visual and RH-sensory information (Fig. 6c and d). By combining these two characteristics, a humidity-adapted visual system is demonstrated based on the MXene-ZnO memristor crossbar array, showcasing noise suppression and filtering during the sensing and preprocessing of visual signals, and it also functions as the artificial synapse for the postprocessing of sensed images. With the RH modulation, good linearity of weight update is achieved, and the system finally exhibits image recognition with a high accuracy of 82.96%.

2.4 Charge trapping/de-trapping *via* defect engineering

Defect engineering is a critical method for tailoring the electrical, optical, or mechanical properties of 2D materials to meet the specific requirements for sensory data processing. By introducing defect states into the 2DM layer through chemical/plasma treatment or laser irradiation, different charge trapping-detraping dynamics can be achieved. These defect states act as charge-trapping sites, enabling the dynamic modulation of the device's optoelectronic properties, such as conductivity and photoresponsivity, with external information stored as trapped charges. By carefully adjusting the external treatment conditions, desirable defect states can be introduced into the materials, achieving precise control of the charge trapping and de-trapping process. This modulation of synaptic strengths facilitates the emulation of learning and adaptation functions in neuromorphic systems.

Liao *et al.* demonstrated an in-sensor visual adaption system using MoS₂ phototransistors, mimicking the dynamic modulation of photosensitivity under different light conditions of the human retina to achieve image contrast enhancement and significantly improve image recognition accuracy.⁴ The visual adaption behavior is achieved by intentionally inducing defect states in MoS₂ phototransistors *via* ultraviolet-ozone (UVO) treatment (Fig. 6e). Ambipolar trap states are created in the bandgap of UVO-treated MoS₂. Under light stimulation, photo-generated carriers are trapped in these defective states, shifting the Fermi level and changing the conductance state (Fig. 6f). Furthermore, an applied back-gate voltage can tune the position of the Fermi level and influence the charge trapping-detraping process at the defect states (Fig. 6g), showing the current excitation/inhibition phenomenon. By employing this

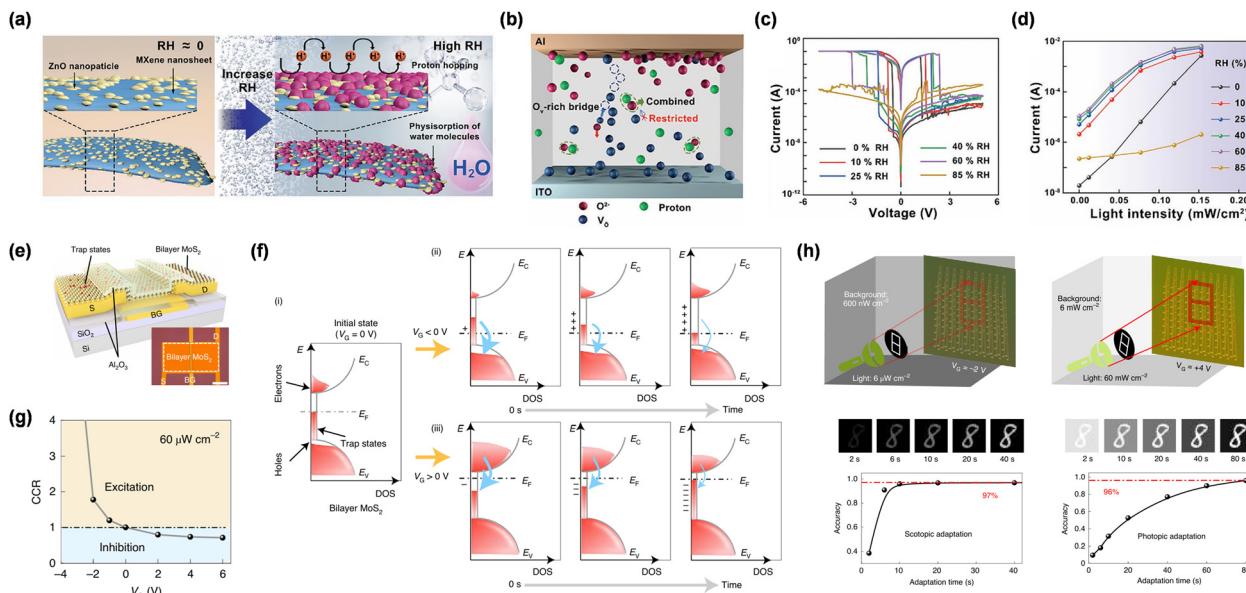


Fig. 6 Multimodal regulation on conductive filament formation. (a) Water molecule's adsorption process at the abundant hydrophilic sites located at the surface of MXene-ZnO heterojunction in high relative humidity (RH) environment. (b) Schematic of the proton-mediated resistive switching mechanism. The combination of protons and oxygen vacancies regulates the conductive filament growth. (c) RH modulation on I – V curves of the MXene-ZnO based memristor. (d) Optical modulation on the current response of the memristor in various RH environments. Reproduced with permission.⁹⁸ Copyright 2021, Wiley-VCH GmbH. Dynamic modulation of photosensitivity by charge trapping process via defect engineering. (e) Schematic of MoS₂ phototransistor with ultraviolet/ozone (UVO)-induced defect states. (f) Schematic of the band structure of the MoS₂ phototransistor under different gate voltages and density of states (DOS). (g) Current change ratio (CCR) of the device at different V_G values, indicating the presence of both excitation and inhibition effects. (h) Scotopic and photopic adaption achieved with efficient photoelectric modulation. Reproduced with permission.⁴ Copyright 2022, Springer Nature Limited.

photoelectric modulation, scotopic and photopic adaptation is successfully achieved by utilizing negative (positive) back gate voltage, V_g , for the dim (bright) background to obtain excitation (inhibition) behavior (Fig. 6h). A $[8 \times 8]$ phototransistor array fabricated using this method exhibited accurate image perception under both dim and bright backgrounds. In addition, by harnessing the conductance change during the charge trapping process, an optoelectric memory was implemented with the MoS₂ phototransistor, enabling the optical information record. Finally, the image recognition task performed using the MoS₂ phototransistor array showed a noticeable improvement in recognition rate exceeding 80% due to its superior photopic adaption.

3. Essential requirements and properties for in-sensor computing

3.1 Evolution of in-sensor computing paradigm

The evolution of in-sensor computing encompasses several phases, marked by the progressive integration of the sensing, processing, and storage units.⁹⁹ Initially, these functions were performed separately in different components, resulting in significant latency due to time-consuming analog-to-digital conversion (ADC) within the sensing unit and frequent data transfer between discrete memory and processor, as seen in the traditional von Neumann architecture. To mitigate intensive

data conversion and transfer, the second phase of development introduced an architecture integrating sensors with in-memory computing units. In this configuration, ADC is incorporated within the system, allowing analog input signals to be directly utilized in the computing process with necessary encoding. Moreover, memory and computing functions are integrated within the same in-memory computing unit, significantly enhancing processing speed and reducing energy consumption. The latest advancements in in-sensor computing systems aim to fully merge sensing and computing units within a single device.¹⁰⁰ Such devices are designed to directly respond to environmental signals, with the response generated during sensing being formatted appropriately for subsequent computing processes. Beyond this integration, state-of-the-art in-sensor computing systems can exhibit responses regulated by multiple sensory signals simultaneously, enhancing feature extraction from multiple aspects and improving the data processing accuracy.

It is evident that data sensing and processing are equally critical within in-sensor computing systems. Due to the diverse requirements of various applications, different processing methods may be employed even for the same type of sensory signals. Consequently, in-sensor computing devices may exhibit various forms of synaptic plasticity, tailored to the specific needs of target applications. Here, we summarize and categorize the most commonly utilized types of plasticity for implementing computing functions within in-sensor computing devices.

3.2 Biomimetic synaptic characteristics for data computation

The implementation of computing capabilities during the information processing in memory hardware is primarily inspired by the learning rules of chemical synapses in biological nervous systems. The connection strength between the presynaptic and postsynaptic neurons is effectively modulated by the release of neurotransmitters, which act as the intermedia to transmit information through the synapse and further regulate the behavior of postsynaptic neurons. Within in-sensor computing devices, the emulation of biological synaptic performance is achieved by the postsynaptic current with continuous adjustments from external stimuli. During this process, the synaptic weights (*i.e.* resistance or conductance) are dynamically controlled by varying input patterns, producing inhibitory (IPSC) or excitatory postsynaptic current (EPSC) in the output terminal.¹⁰¹ Depending on the intrinsic characteristics of different devices, synaptic connections change in different manners, resulting in the creation of various synaptic plasticity that are suitable for a wide range of applications.

Short-term plasticity: to meet the demands of capturing temporal information in rapidly changed environments (light, sound, or smell), short-term plasticity (STP) is favorable for real-time signal processing. It refers to the temporary changes

in synaptic strength that occur on timestep ranging from milliseconds to minutes and its transient nature enables rapid and reversible adjustments in synaptic efficacy. STP contains both short-term facilitation and depression, which are always characterized by paired-pulse facilitation that investigates the response of the device under the stimuli of two identical consecutive pulses with a small interval τ . Normally the PSC shows a more significant increase when a smaller interval is adopted and the time-interval-dependent PSC gain is evaluated by the PPF index, which is determined by the following equation:

$$\text{PPF index} = \frac{A_2}{A_1} = \frac{A_2 - A_1}{A_1} = C_1 \cdot \exp\left(-\frac{\Delta t}{\tau_1}\right) + C_2 \cdot \exp\left(-\frac{\Delta t}{\tau_2}\right)$$

where A_1 and A_2 denote the amplitude of the device response to the first and second stimulation. τ_1 (C_1), and τ_2 (C_2) represent the characteristic relaxation time (initial facilitation magnitude) corresponding to the rapid and slow decay process, respectively.^{101,102}

In a common retina-inspired photodetector, Duong *et al.* demonstrated STP in an α -In₂Se₃-based retinomorphic sensor. Under the co-modulation of electrical and optical stimuli, the device exhibits gate-tunable excitatory and inhibitory functions in photon-induced short-term plasticity.⁵ The laser-induced paired-pulse facilitation is measured as shown in Fig. 7a, with

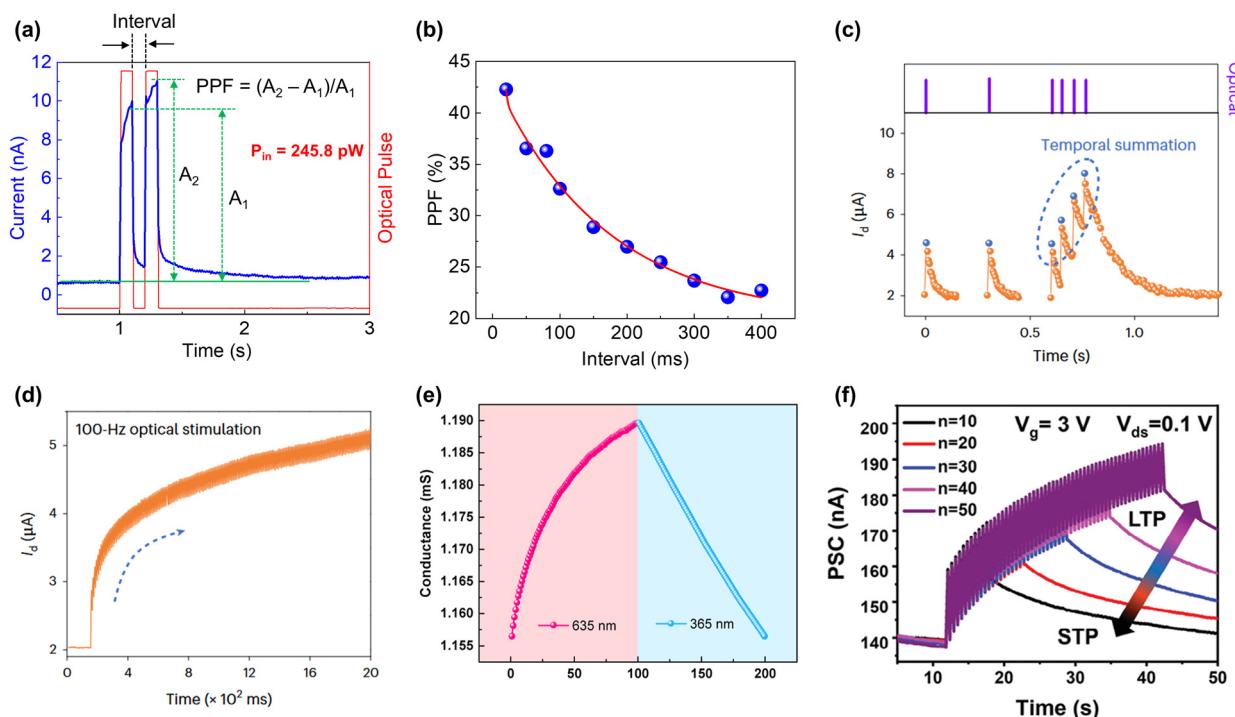


Fig. 7 Biomimetic synaptic characteristics for data computation. (a) Characterization of short-term plasticity using paired-pulse facilitation. (b) Extracted PPF ratio corresponding to measurement results in (a), as a function of the pulse interval. Short-term plasticity in MoS₂-based phototransistor. Reproduced under the terms of Creative Commons CC BY License.⁵ Copyright 2023, The Authors. Advanced Science published by Wiley-VCH GmbH. (c) Drain current exhibits a dramatic increase under stimulation and gradually decays towards zero once the light is removed, showing temporal summation behavior. (d) Distinguishable device states obtained from short-term potentiation. Reproduced with permission.¹⁰³ Copyright 2022, Springer Nature Limited. (e) Realization of synaptic conductance modulation with all optically controlled modulation in a graphene/titanium dioxide quantum dot heterostructure. Reproduced with permission.¹⁰⁴ Copyright 2023, American Chemical Society. (f) The transition from STP to LTP induced by increasing the duration of light pulses. Reproduced under the terms of Creative Commons CC BY License.⁸⁹ Copyright 2023 The Authors. Advanced Science published by Wiley-VCH GmbH.

extracted PPF ratio that changes along with varying pulse intervals (Fig. 7b). The PPF measurement verifies the sub-millisecond fading dynamics of the device, which is vital for the implementation of dynamic kernels for convolutional image processing. In addition, as shown in Fig. 7c and d, Chen *et al.* demonstrated temporal summation characteristics in MoS₂-based phototransistors.¹⁰³ This STP assists the encoding process of light stimuli in the temporal domain and contributes to the construction of bioinspired vision sensors that emulate the graded neuron inside the insect visual system.

Long-term plasticity (LTP) and plasticity transition: different from the short-term plasticity that is associated with adaptive response and spatiotemporal information processing, long-term plasticity is preferably related to long-term learning algorithms (e.g. Hebbian learning rule) and memory capabilities.¹⁰⁵ The implementation of LTP in a synaptic device not only requires effective modulation of connect strength directly by external stimuli but also needs to make sure the obtained modifications are nonvolatile. In experiments, the LTP is usually characterized by investigating the existence of long-term potentiation and depression behaviors. Linearity and symmetry are two important parameters to evaluate the LTP within a device. To realize a promising LTP, good linearity is expected under the stimuli of sequential identical pulses. Although nonlinearity can be improved by using a pulse sequence with progressively enhanced pulse amplitude or prolonged pulse interval, this comes at the expense of introducing complex peripheral circuits inside the system.

In-sensor computing devices extend beyond simple modulation by electrical stimuli; they can achieve long-term potentiation and depression through the synergistic regulation of diverse external signals. As shown in Fig. 5j, co-modulation of both electrical and optical stimuli is a widely adopted method for synaptic conductance modulation in optical synapses, including an optical stimuli-controlled LTP process and an electrical stimuli-controlled LTD process. However, this mixed-input control for long-term plasticity usually relies on complex external circuitry to generate the necessary pulse schemes, leading to great latency in overall processing times. An alternative approach, which shows greater promise, employs all-optical regulation for synaptic conductance modulation.^{104,106} As depicted in Fig. 7e, this method allows bidirectional synaptic conductance modulation using solely optical stimuli, thereby simplifying the system complexity and accelerating processing time.

Additionally, synaptic responses in the form of STP and LTP are not discrete. A gradual transition from STP to LTP can be achieved through repeated stimulation, mimicking the memory consolidation process in the human brain where information transfers from the hippocampus to the cerebral cortex. Wang *et al.* demonstrated a CIPS-based optoelectronic FeFET in which the transition from STP to LTP is achieved by adjusting the pulse count or duration (Fig. 7f).⁸⁹ This mechanism enables the device to display both volatile and nonvolatile characteristics, facilitating seamless transitions between these states and broadening the applicability of in-sensor computing devices.

4. Information processing via 2DM in-sensor computing

According to the preceding discussion, 2DM-based optoelectronic devices exhibit significant potential for processing sensory information by leveraging various physical mechanisms such as ferroelectricity and charge trapping/de-trapping. These devices are highly suitable for hardware implementation of in-sensor computing systems due to their superior material properties and high scalability. Unlike conventional artificial sensory systems that require considerable energy and exhibit high latency due to redundant sensory data transmission between computation units and sensors, in-sensor computing eliminates the physical separation between sensing, memory, and processing units. In this integrated system, sensory devices perceive external information and perform computational tasks, allowing parallel information processing directly within the sensors. This fusion of perception and computation within a single device substantially enhances processing speed and energy efficiency. Recent reports have highlighted various 2DM-based in-sensor computing systems that achieve efficient sensory data processing, surpassing the performance of traditional in-sensor computing systems. Additionally, integrating these sensory devices with innovative system architecture design, such as parallel and grouped physical reservoirs, further enhances the feature extraction capability and overall performance of in-sensor computing systems.^{107,108}

4.1 In-sensor reservoir computing system

A notable architecture within in-sensor/near computing systems is in-sensor reservoir computing. Reservoir computing, refined from the recurrent neural network, is inherently suitable for temporal data processing due to its reservoir layer with numerous cyclically connected nodes. This structure projects inputs into high-dimensional space, extracting various features.¹⁰⁹ In addition, the fixed weights between the input and reservoir layers necessitate simple training for the readout layer, reducing energy consumption and processing speed.^{110,111} To efficiently implement a physical reservoir, two key properties are essential: nonlinearity and short-term plasticity. Nonlinearity ensures the input signals are projected into a higher dimensional space, making previously inseparable inputs linearly separable due to increased feature differentiation.¹¹² Short-term plasticity is crucial for realizing cyclic connections within the reservoir, ensuring reservoir states depend not only on current inputs but also on recent past inputs.¹¹³ Sensory devices with these properties can implement in-sensor reservoir computing systems, integrating computation and sensory functions to handle temporal sensory data effectively.

Among various approaches, memristor-based RC systems have gained the most attention owing to their simple fabrication and integration. However, the relatively fixed nonlinear transformation function of memristors makes it difficult to adjust timescales, thus limiting their performance.^{107,114,115} To address this, a multimode and multiscale reservoir computing system was developed using 2D α -In₂Se₃-based memristors

with a structure of Au/Pd/ α -In₂Se₃/Pd/Au (Fig. 8a).¹¹⁶ Leveraging the in-plane ferroelectricity and photoresponsivity of α -In₂Se₃, the memristor exhibits tightly coupled electrical and optical synaptic plasticity, allowing for tunable timescales modulated by this hetero-synaptic plasticity. Electric-field-driven ferroelectric switching results in noticeable potentiation and depression behaviors under electrical pulse train stimuli, with higher stimulus frequencies leading to more pronounced potentiation or depression.

Similarly, optical pulses induce an accumulation effect modulated by light intensity and frequency due to photo-generated carriers (Fig. 8b). These two modulation types interact, generating diverse current levels (reservoir states). For electrical stimulation, the relaxation time of inhibitory post-synaptic current (PSC) decreases under constant light, and similar modulation can be achieved with light under negative back-gate voltage conditions

(Fig. 8c). This tunable fading memory behavior and the coexistence of optoelectrical modulation enable a mixed-input RC system that demonstrates multi-feature extraction and multisensory fusion with one physical reservoir (Fig. 8d). Unlike conventional RC systems, which require multiple reservoirs to process various inputs individually before combining them for classification, this α -In₂Se₃-based multimode RC system dramatically reduces hardware costs and energy consumption. The system's superiority is demonstrated with a multimode hand-written digit recognition task, achieving an 86.1% accuracy that outperforms the performance of the RC system that can only handle one type of input (Fig. 8e and f). Moreover, the prediction of multiple superimposed oscillator (MSO) time-series data containing various frequency information is successfully demonstrated with a low normalized-root-mean-square error of 0.105, further proving the capability of

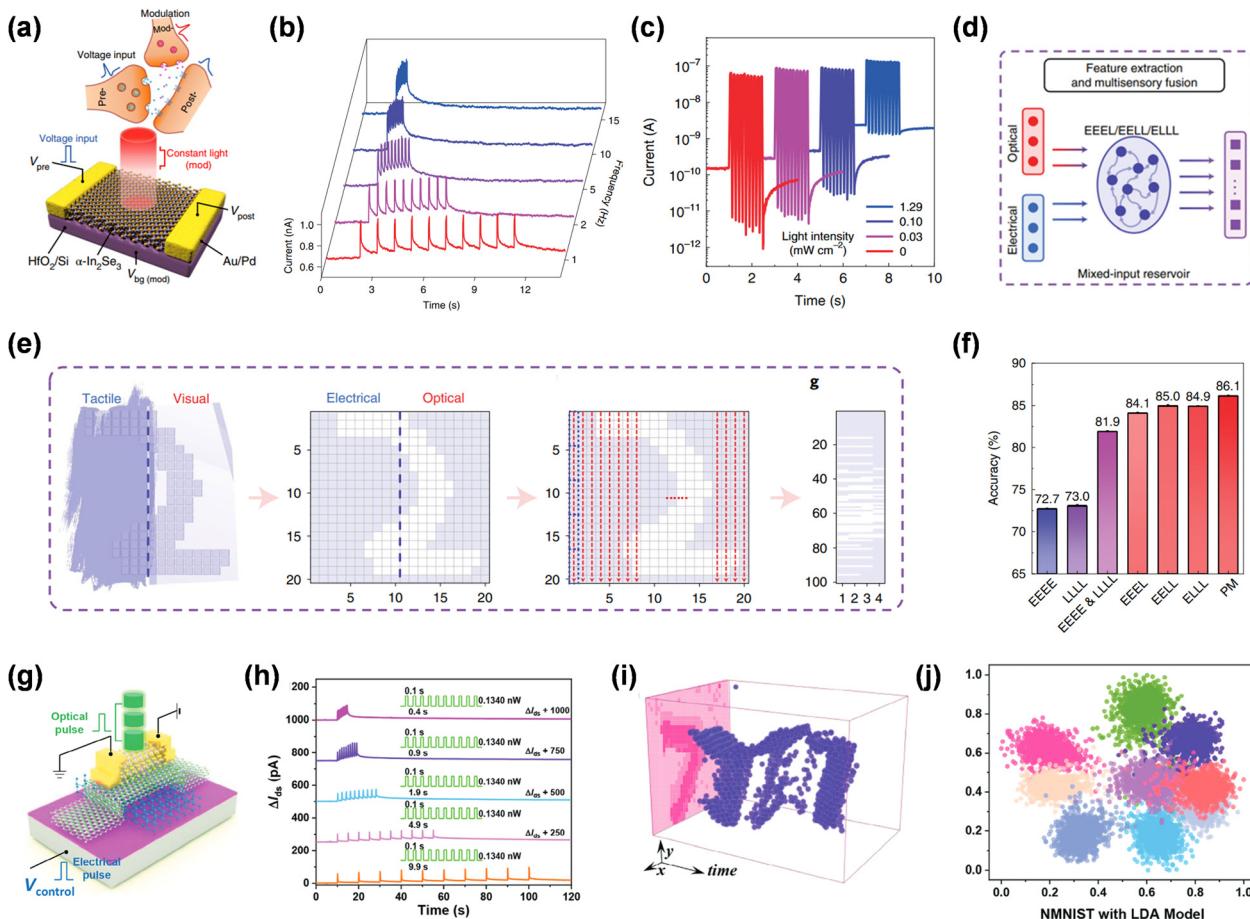


Fig. 8 In-sensor reservoir computing systems. (a) Schematic of α -In₂Se₃-based optoelectronic heterosynapse with additional light/back-gate modulating terminal. (b) Tunable excitatory post-synaptic current (PSC) induced a train of optical pulses. (c) Light modulation on electrical pulse induced inhibitory PSC, demonstrating tunable relaxation time and synaptic behavior of the device. (d) Demonstration of mixed-input reservoir computing based on multi-sensory fusion of the α -In₂Se₃-based optoelectronic for feature extraction. (e) Recognition of images composed of mixed types of inputs (both tactile and visual inputs). (f) Multi-mode digit recognition accuracy at different working modes. Reproduced with permission.¹¹⁶ Copyright 2022, Springer Nature Limited. (g) Schematic of MoS₂/h-BN/Te heterostructure-based optoelectronic FG memory unit. (h) Various volatile optoelectronic memory behaviors are obtained with optical pulse modulation, showing rich fading memory dynamics. (i) A sample from the N-MNIST dataset. The event stream of the N-MNIST dataset possesses a higher spatial resolution (34×34 pixels) than the MNIST dataset. (j) Featurizing encoding by the multimodal reservoir using linear discriminant analysis (LDA), exhibiting effective nonlinear transformation and dimensionality reduction. Reproduced with permission.¹¹⁷ Copyright 2023, Wiley-VCH GmbH.

the α -In₂Se₃-based in-sensor RC system to handle sensory information.

Despite the satisfactory performance of the 2DM memristor-based in-sensor RC system in sensory data processing, challenges remain. Typically, the conductance states are used as reservoir states in memristor-based RC systems, necessitating a read pulse after each programming pulse to read out the conductance state, limiting processing speed and increasing energy consumption.^{116,118} In addition, the difficulty of addressing single device nodes within the crossbar arrays and minimizing the sneak current also caused another practical challenge in the memristor-based RC system. Sneak path currents can lead to unintended currents flowing through other memristors in the array that are not directly accessed and cause additional leakage currents, resulting in the wrong reading and writing of data and increasing the overall energy consumption. Consequently, transistor-based RC systems have gained more attention due to their better control abilities with an additional gate terminal and the absence of sneak path current.

As demonstrated in Fig. 8g, a 2DM heterostructure-based floating gate memory device leverages gate modulation to control the charge-trapping process, functioning as both memory and an in-sensor physical reservoir.¹¹⁷ The MoS₂/hexagonal boron nitride (h-BN)/Te heterostructure is formed by drop-casting solution synthesized Te nanoflakes and dry transfer of exfoliated h-BN and MoS₂ flakes. The 2D Te serves well as a floating gate due to its high trapped charge density and rapid recombination at shallow trap states. Dual-type input stimuli (electrical and optical) can both rely on the charge-trapping and neutralization process to achieve nonvolatile and volatile behavior. Electrical pulses applied to the control gate modulate electron tunneling from MoS₂ to TE, and the trapped charge in the FG caused the change of device states. As for optical stimuli, they generate photo-excited carriers that can neutralize the stored electrons inside the FG. In addition, it is easy to switch between the nonvolatile and volatile behavior by simply tuning the stimuli intensity. A superior and stable nonvolatile behavior is obtained under high-intensity stimuli, while weak stimuli provide volatile behavior with rich dynamics (Fig. 8h). The performance of this heterostructure-based RC is examined using the N-MNIST composed of temporal information and changes in pixel intensities rather than the static image in the MNIST dataset (Fig. 8i). By leveraging the multisensory fusion of the heterostructure-based RC system, effective nonlinear transformation and dimensionality reduction are achieved during feature extraction process (Fig. 8j), resulting in a 90.77% accuracy, significantly higher than the single-mode RC system and surpassing the software baseline.

4.2 In-sensor motion detection

Visual information constitutes a significant portion of the vast data generated in the current era of the internet-of-things. Applications such as autonomous vehicle safety and surveillance systems require efficient motion processing systems. However, traditional hardware implementation of vision sensors relies on complex algorithms requiring iterative solutions, leading

to high computational costs and large footprints. Additionally, capturing visual information generates large amounts of redundant data, necessitating the transmission of computing units for processing. The physical separation between vision sensors and processing units introduces delays and increases power consumption. Many image sensors can only output spatial frames without encoding the temporal information, limiting their use to static image processing and presenting significant challenges for dynamic motion processing.¹¹⁹ Inspired by biological visual systems that efficiently detect motion in complex environments with minimal computational resources, the emerging vision sensors based on in-sensor computing schemes offer a promising solution. By integrating computation functions into the sensory devices, these sensors avoid additional analog-digital conversion and unnecessary data movements.¹²⁰ Recently, many 2DMs have been utilized to design efficient vision sensors due to their extraordinary photosensitivity, and 2DM-based vision sensors have demonstrated outstanding motion processing capabilities.

In insects, the lobula giant movement detector (LGMD) neurons, located in the lobula of their brains, are sensitive to the expansion of objects in the visual field and play a crucial role in collision avoidance behavior.^{121,122} Inspired by LGMD neurons, a collision avoidance system for autonomous vehicles is designed (Fig. 9a). However, traditional implementations require very-large-scale-integration (VLSI) systems, which are energy and area inefficient.^{123–125} To overcome these bottlenecks, a nanoscale collision detector was designed by stacking a monolayer MoS₂-based photodetector on top of a non-volatile floating-gate (FG) memory (Fig. 9b).¹²⁶ This detector effectively performs collision detection while consuming minimal energy (nanojoules) and having a small device footprint ($\sim 1 \mu\text{m} \times 5 \mu\text{m}$). It successfully mimics the escape response of LGMD neurons by exhibiting an excitatory response to light stimulation (Fig. 9c) and an inhibitory response to electrical stimulation (Fig. 9d). During collision detection, a static light source with changing intensity (*i.e.*, looming stimuli) simulates an approaching car. Owing to the photoexcitation effect of the MoS₂ detector, the device exhibits a monotonic current increase when exposed to the light source. The inhibitory behavior is a result of dynamic engineering of the threshold voltage caused by charge trapping between the back gate and the floating gate under a programming voltage. Integrating these two competing mechanisms, the device shows an inflection point before the collision occurs (when the light source reaches maximum intensity), indicating its capability to work as a collision indicator. The MoS₂-based collision detector demonstrates high energy and area efficiency as it comprises only a single device without requiring an array or peripheral circuits. It also works effectively with both direct incident light from a moving object and reflected passive light from the collision object, demonstrating its versatile functionality (Fig. 9e).

Moreover, the intrinsic defects in the as-grown MoS₂ film can be employed to implement learning rules. By coupling photoexcitation with shallow trap centers in MoS₂, an optoelectronic phototransistor was successfully implemented to

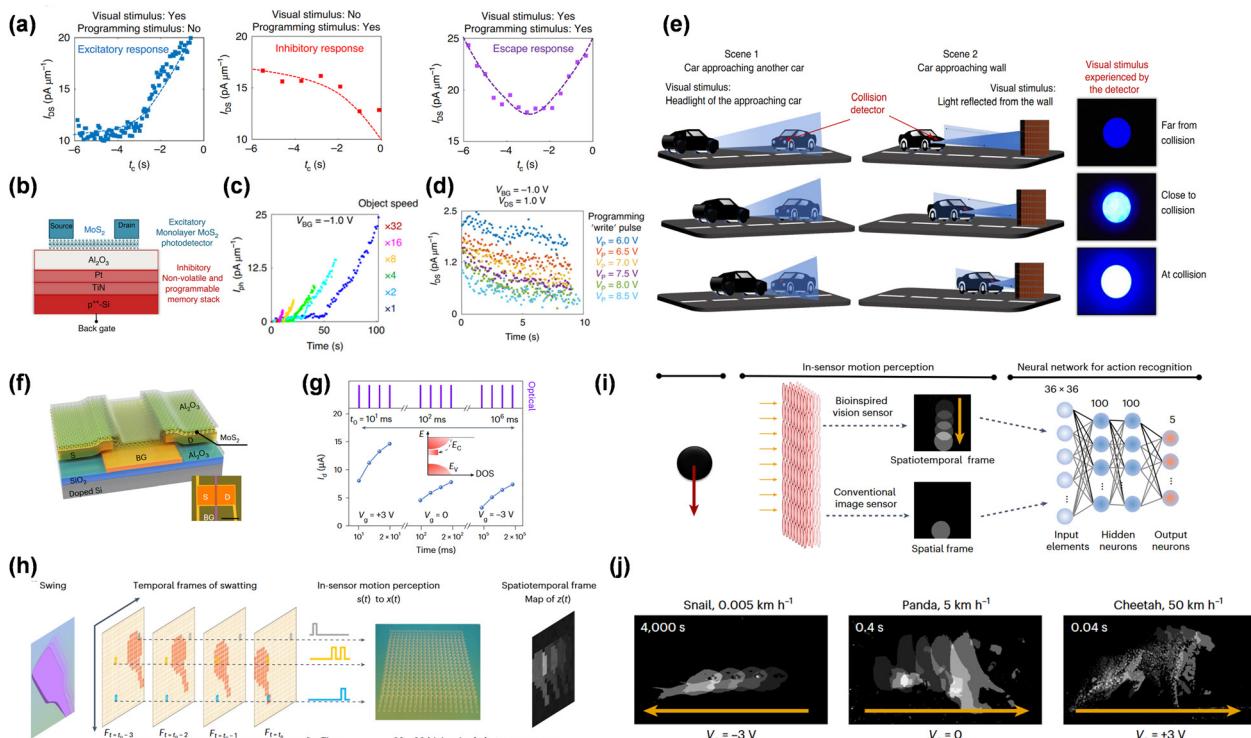


Fig. 9 In-sensor motion-detection systems. (a) Optoelectronic collision detector mimicking the working principles of LGMD neurons. Excitatory, inhibitory, and escape responses are obtained when only looming optical stimulus, only electrical stimulus, both optical and electrical stimulus applied, respectively. (b) Schematic of a biomimetic collision detector consisting of a MoS₂-based photodetector and a programmable FG non-volatile memory stack. (c) Excitatory photocurrent responses in the MoS₂ FET at negative constant back-gate biasing. (d) Inhibitory output current obtained with varying electrical programming pulses. (e) Schematic of two different collision detection scenarios emulated using LED light source: (1) direct light from collision object (2) passive light reflected from collision object. Reproduced with permission.¹²⁶ Copyright 2020, Springer Nature Limited. (f) Schematic of the MoS₂ phototransistor-based vision sensor. (g) Gate voltage modulation on the position of defect states. The device exhibits distinct temporal resolution, ranging from 10 ms to 10⁶ ms. (h) Spatiotemporal information encoding using MoS₂ phototransistor array, accurately depicting the trajectory contour of the moving object. (i) Perception of motion direction using bioinspired vision sensor with output containing spatiotemporal frames. (j) Successful spatiotemporal information encoding across varying timescales. Reproduced with permission.¹⁰³ Copyright 2023, Springer Nature Limited.

perceive different types of motions (Fig. 9f).¹⁰³ Under light stimulation, photogenerated carrier trapping into existing defects produces a nonlinear conductance increase, and the trapped charges are released spontaneously upon light removal, emulating the function of graded neurons. Furthermore, effective gate voltage modulation tunes the charge dynamics of shallow trapping centers, enabling various temporal resolutions ranging from 10 to 10⁶ ms (Fig. 9g). This efficient encoding of temporal information fuses spatiotemporal features across varying timescales into compressive images (Fig. 9h), achieving a high recognition accuracy of 99.2% for direction perception (Fig. 9i). This setup allows for the detection of motion with varying speeds and enhanced vision saliency of moving objects (Fig. 9j), significantly outperforming the conventional image sensors.

4.3 Multimodal sensing and processing

While visual information processing has dominated in-sensor computing research, it is imperative to explore the potential of these devices in handling other types of sensory information. Pressure detection, for instance, is essential in various applications, including tactile sensing in robotics,¹²⁷ wearable health

monitors,¹²⁸ and smart electronics.¹²⁹ In-sensor computing devices leveraging 2D material transistors and memristors offer promising advancements in this field. These materials provide excellent mechanical flexibility, high sensitivity, and low power consumption, making them ideal for developing responsive and accurate pressure sensors. By integrating 2D materials such as graphene, MXenes, and transition metal dichalcogenides (TMDs), researchers are creating innovative pressure detection systems that can seamlessly process sensory data at the point of collection. This section delves into the latest developments and applications of in-sensor computing devices for various sensory data sources, including pressure detection and gustatory information, highlighting how these technologies expand the horizons of sensory information processing beyond the visual spectrum.

The demonstration of a high-performance tactile information processing system composed of seamlessly connected tactile sensors and memristor arrays has been realized based on MXene, showcasing the potential for Morse code recognition and human health state monitoring.¹³⁰ Requiring no high-temperature processes or complex material growth or transfer processes, the sensor and memristor are integrated, utilizing

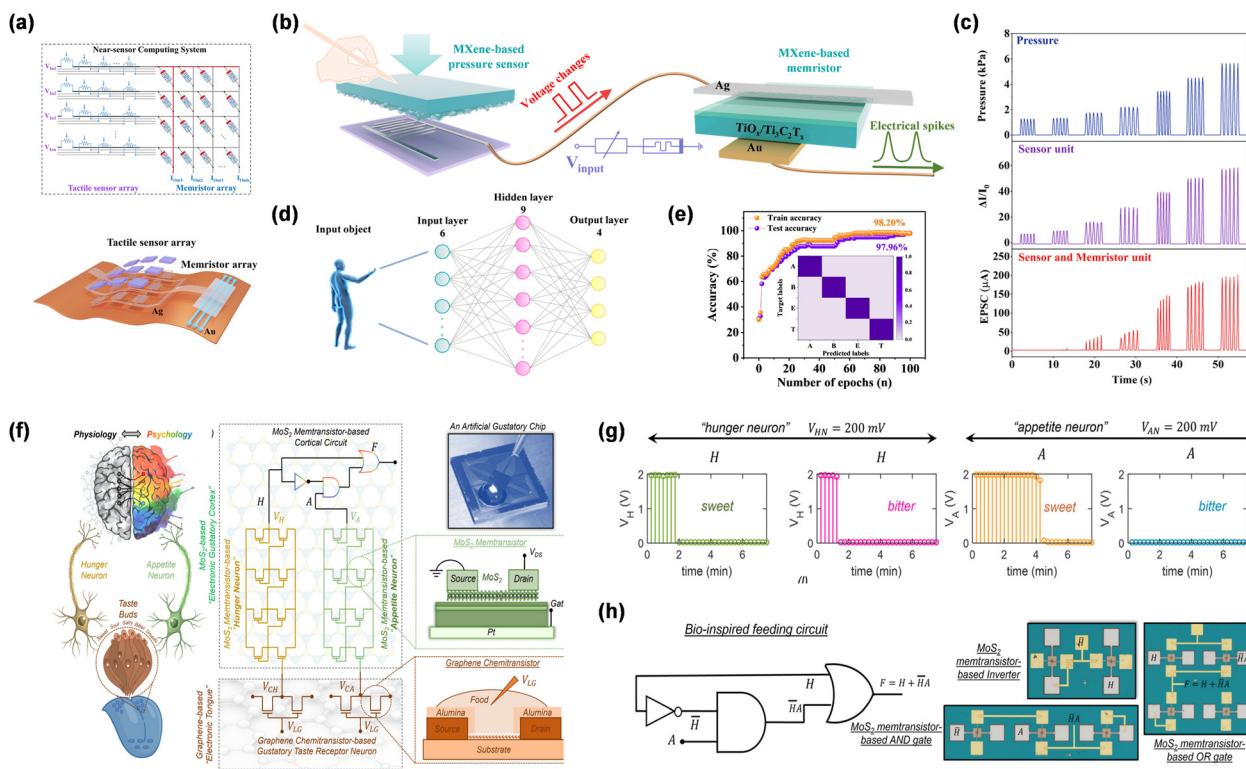


Fig. 10 Multi-modal in-sensor computing schemes. (a) Schematic of near sensor computing system constructed using MXene-based sensory neuron array and memristor array. (b) Signal transmission processing between the sensor unit and synapse unit. (c) Conversion between tactile signals and postsynaptic current within the sensory neuron. (d) Schematic three-layer fully connected ANN constructed based on MXene-based flexible sensory neuron for human respiratory state classification. (e) Confusion matrix and recognition accuracy for the human respiratory state classification task. Reproduced with permission.¹³⁰ Copyright 2024, Elsevier Ltd. (f) Schematic of the bioinspired gustatory system consisting of graphene chemitransistor-based sensing unit monolayer MoS₂ memtransistor-based computing unit. (g) Evolution of temporal response to sweet and bitter taste stimulants in hunger and appetite neurons. (h) Schematic of adaptive feeding circuit using MoS₂-based memtransistor as building blocks. Reproduced under the terms of Creative Commons CC BY License.³⁶ Copyright 2023, The Author(s). Published by Springer Nature.

only a single functional MXene layer (Fig. 10a), significantly reducing the transmission delay and energy cost and simplifying the fabrication process. Due to the micropump structure produced by sandpaper templates in PDMS films, under certain physical pressure, the contact area between the MXene sensitive layer and PDMS films keeps changing and converts the physical pressing signal into action potentials. Then, the generated stimulation is transferred into the MXene-based memristor array for processing (Fig. 10b). The single functional MXene layer enables efficient sense and conversion of external tactile signals (Fig. 10c) and shows highly linear and symmetric weight updates in the synaptic array, implying its capability for information encoding and processing. A high accuracy of 97.96% is achieved for human respiration state classification tasks (see Fig. 10d and e).

Besides the sensing and processing of physiological behavior, novel bio-inspired 2DM-based in-sensor computing systems also exhibited the capability to mimic the psychological behavior of humans. An all-2D gustatory circuit mimicking the physiology and psychology of adaptive feeding behavior in humans has been proposed.³⁶ As demonstrated in Fig. 10h, this innovative system integrates graphene-based chemo-transistors as artificial

gustatory taste receptors, forming an electronic tongue that allows the device to detect and process taste information, together with monolayer MoS₂ memtransistors constructing an electronic gustatory cortex, enabling the system to simulate the neural processes involved in feeding behavior (Fig. 10i). The advantages of different 2DMs are fully harnessed for different implementations: due to the high carrier mobility at room temperature,¹³¹ inherently low electrical noise,¹³² and enhanced sensitivity caused by the exceptionally high surface-to-volume ratio,¹³³ graphene is adopted for artificial taste receptors. MoS₂ is utilized to emulate cortical decision function due to its superior semiconducting properties, which are beneficial for signal processing and decision-making. Such in-sensor computing device not only capable of co-processing physical and physiologic information, but also establishes a novel paradigm for emotional-AI systems that can bridge the gap between human and machine intelligence, with significant implications for human health and future AI applications.

In addition to processing tactile and gustatory information, 2DM-based sensors have also exhibited superior capabilities in dealing with various sensory sources (e.g., gas^{7,134,135} or audio information^{77,128}). Integrating 2DM with intelligent memristors and transistors overcomes the limitation of in-sensor

computing devices that rely solely on electrical modulation, significantly expanding their application in perception computing.

5. Conclusion and future outlooks

In-sensor computing architecture demonstrates significant potential and efficiency in processing sensory signals with complex temporal information and high data intensity. It quickly responds to various types of sensory data and provides accurate and energy-efficient processing of captured information. The implementation of in-sensor computing systems using two-dimensional materials further underscores their superiority because of the superior material properties of 2DMs and the novel device architectures derived from 2DM heterostructures. This review presents an overview of the development of in-sensor computing architecture, detailing the mechanisms inherent to 2DMs that can be harnessed for such systems and the necessary synaptic functions for effective data computation within the devices. Additionally, the potential applications of 2DM-based in-sensor computing systems are discussed and categorized. A comparison of state-of-the-art 2DM-based in-sensor computing systems is summarized in Table 1.

5.1 Challenges and opportunities

Artificial intelligent sensors based on emerging 2D semiconductor devices offer low power consumption, reduced circuit complexity, and novel functionalities. However, several significant challenges must be addressed to enable large-scale integration of these devices for commercial vision sensors.

Multiple sensory computing: a fundamental requirement for multimodal sensory computing is the development of advanced sensors capable of capturing various types of sensory data. Although there has been notable progress in developing sensors for individual modalities such as vision, touch, and olfaction, the ability to effectively capture and process multimodal data in real time remains a significant challenge. A key question in multisensory integration is how the brain determines the weighting of each sensory input when forming a combined percept. One prominent theory is the “statistical optimization model” or “Bayesian integration model,” which suggests that the brain weights each sensory cue based on its reliability derived from its prior probability. Integrating data from multiple sensors due to differences in temporal and spatial resolutions, noise characteristics, and units of measurement. Developing efficient algorithms and hardware architectures to fuse this heterogeneous data into a coherent representation is crucial for enabling meaningful multimodal perception and decision-making.

Integration scale: large-scale integration of sensing-memory-computing devices into arrays presents a significant hurdle. Current vision chips typically feature a relatively small scale of 1000 to 100 000 pixels with a substantial device-to-device variation, impeding large-scale integration and limiting their suitability for practical applications.¹⁴⁵ Furthermore, the absence of a compact physical model for emerging vision chip

technologies based on 2D semiconductors complicates the prediction of device performance in larger systems.

Material system and device structure trade-offs: developing materials and device structures suitable for multifunctional sensing-memory-computing electronics poses a significant challenge. The ideal material must perform stably across sensing, memory, and computing functions. For example, proposing memristors and light-induced ferroelectric materials with considerable photo-response across a wide light range is crucial for artificial visual systems. In-sensor computing also necessitates the development of novel device functions and mechanisms. While direct analogy processing offers real-time readout and reduced data conversion, it is vulnerable to noise and complex to design on a large scale, requiring optimization for specific applications. Developing new algorithms will be necessary for non-visual data, as features from auditory and olfactory sensors have lower dimensionality than visual signals.

Ferroelectric materials: ferroelectric materials offer low-power operation and air stability for sensing-memory-computing devices, making them potentially suitable for retina-inspired sensors. For instance, $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$ exhibits pronounced ferroelectric properties, displaying robust polarization at nanoscale thicknesses. However, interface states, ambiguous mechanisms, and integration challenges hinder the advancement of ferroelectric materials for in-sensor computing. Further research is essential to elucidate the material properties and enhance polarization characteristics, thereby facilitating the development of high-performance semiconductor devices.

5.2 Future outlooks

5.2.1 Key challenges. The development of 2D semiconductor-based in-sensor computing hardware necessitates interdisciplinary collaboration across various domains, including sensors, devices, integration technologies, and algorithms. The future of in-sensor computing aims to address several key challenges:

Multimodal processing: current research predominantly focuses on single-sensory processing. Future intelligent devices and systems must be capable of real-time fusion of multiple sensory nodes, such as visual, auditory, olfactory, and tactile. This advancement is crucial for applications in robotics, intelligent vehicles, and wearable electronics. The ultimate objective is to develop intelligent systems that can perceive and interact with the environment in a human-like manner.

The successful development of multimodal sensory computing technologies could revolutionize various fields. In robotics, multimodal sensor systems could enhance robots’ perception capabilities, enabling them to navigate complex environments, interact safely with humans, and perform intricate tasks. In autonomous driving, combining data from cameras, lidar, radar, and other sensors could allow vehicles to perceive their surroundings more accurately and reliably, improving safety and driving capabilities. Wearable electronics for healthcare could also benefit, with multimodal sensors integrated into devices like smartwatches and fitness trackers providing comprehensive health monitoring, personalized feedback, and

Table 1 Benchmarking of state-of-the-art 2DM-based in-sensor computing systems

Device structure	Functional 2DM	Sensor source	Mechanism	Application	Performance	Ref.
(ITO)/MXene/ZnO/Al/PDMS memristor	MXene	Relative humidity	Multi-field controlled conductive filament formation	Noise suppression & image recognition	82.96%	98
Ag/Mxene/Au/PI memristor	MXene	Pressure	Conductive filament formation	Human state classification	97.96%	130
Cr/Au/MoS ₂ /SiO ₂ /Si phototransistor	MoS ₂	Light	Light-modulated charge trapping/detrapping	Handwritten digital recognition	96.90%	4
Pt/Au/WSe ₂ /h-BN/Ti/Au/Si/SiO ₂ transistor	WSe ₂ /h-BN	Light	Charge trapping & photoresponsivity	Color-mixed pattern recognition	Over 90%	23
Cr/Au/SnS/Si/SiO ₂ memristor	SnS	Light	Charge trapping & photogating effects	Practical Korean sentence classification	91%	118
Cr/Au/MoS ₂ /h-BN/Te/SiO ₂ /Si transistor	MoS ₂ /h-BN/Te	Light	Charge trapping & photoresponsivity	Event-based MNIST digit-recognition	90.77%	117
Cr/Au/MoS ₂ /Al ₂ O ₃ /Cr/Au/SiO ₂ /Si phototransistor	MoS ₂	Light	Photoexcitation & gate-tunable charge trapping	Moving ball track recognition	99.2%	103
SiO ₂ /Si phototransistor	h-BN/CIPS	Light	Light-modulated ferroelectric polarization	(MNIST) handwritten digital recognition	93.62%	89
SnS ₂ /h-BN/CIPS/Au/SiO ₂ /Si FeFET	CIPS	Pressure	Ferroelectric-enhanced piezoelectricity	Speech recognition & motion monitoring	96%	128
Cr/Au/CIPS/PET	Graphite/CIPS	Light	Reconfigurable photovoltaic effect	Pattern recognition	Around 95%	136
Graphite/CIPS/graphite/Si/SiO ₂ memristor	α -In ₂ Se ₃	Light	Ferroelectric polarization & photoresponsivity	Handwritten digital recognition	95%	5
Ni/ α -In ₂ Se ₃ /HfO ₂ /SiO ₂ /Si	α -In ₂ Se ₃	Light	Ferroelectricity regulated Schottky barrier height modulation	Pattern recognition	80%-97%	90
Au/graphene/ α -In ₂ Se ₃ /graphene/SiO ₂ /Si	α -In ₂ Se ₃	Light	Photoinduced ferroelectric polarization switching	Visual perceptual processing	N.A.	137
Pd/Au/ α -In ₂ Se ₃ /Ti/Au phototransistor	α -In ₂ Se ₃	Light	Coupled ferroelectric & optoelectronic effects	Multimode handwritten digit recognition	86.1%	116
Pd/Au/ α -In ₂ Se ₃ /h-BN/CIPS/Ti/Au	α -In ₂ Se ₃ /h-BN/CIPS	Light	Electrical and optical modulation	Optical logic operations & handwritten digital recognition	92.5%	91
Cr/Au/graphene/MoTe ₂ /PVDF/Si photodiode	Graphene/MoTe ₂	Light	Ferroelectric-tuned photoresponsivity	Pattern recognition & classification on robot dog movement	N.A.	138
Cr/Au/PdSe ₂ /MoTe ₂ /SiO ₂ /Si photodiode	PdSe ₂ /MoTe ₂	Light	Gate-tunable broadband photovoltaic effect	Broadband image classification	Nearly 100%	22
Ti/Au/graphene/graphene/Ge/SiO ₂ /Si Ti/Au/WSe ₂ /Al ₂ O ₃ /Ti/Au photodiode	Graphite/graphene WSe ₂	Light	Gate-tunable photovoltaic effect	Human face recognition	Over 90%	63
Cr/Au/h-BN/b-AsP/MoTe ₂ /SiO ₂ /Si	b-AsP/MoTe ₂	Light	Photovoltaic & photothermoelectric effects	Image encoding & handwritten digit recognition	N.A.	13
Au/MoS ₂ /HfO ₂ /Ti/Au/PEFT (SiO ₂ /Si) phototransistor	MoS ₂	Light	Gate-tunable photoresponsivity	Handwritten digit recognition	96%	139
Au/h-BN/WSe ₂ /Al ₂ O ₃ /BP/Al ₂ O ₃ /Si transistor	h-BN/WSe ₂ /BP	Light	Gate-tunable photoresponsivity	Handwritten digit recognition	90.64%	10
Pd/Au/WSe ₂ /h-BN/Al ₂ O ₃ /Ti/Au	WSe ₂	Light	Gate-tunable photoresponsivity	Detection and recognition of moving trolleys	100%	11
Cr/Au/h-BN/WSe ₂ /Cr/Pd/Si/SiO ₂ photodiode	WSe ₂	Light	Gate-tunable photoresponsivity	Image stylization, edge enhancement, contrast correction	N.A.	15
Au/WSe ₂ /Al ₂ O ₃ /HfO ₂ /Al ₂ O ₃ /Cr/Au/SiO ₂ /Si Cr/Au/MoS ₂ /sr-SiN _x /Si phototransistor	WSe ₂ /MoS ₂	Light	Programmable photoresponsivity	Convolutional image processing	N.A.	140
Au/h-BN/MoS ₂ /graphene/Si/SiO ₂ transistor	hBN/MoS ₂ /graphene	Light	Photoconductivity	Motion recognition	92%	141
Cr/Au/WSe ₂ /Al ₂ O ₃ /Cr/Au phototransistor	WSe ₂	Light	Photocoupling effect	All-day motion detection and recognition	N.A.	142
Ni/Au/MoS ₂ /Al ₂ O ₃ /Pt/TiN	MoS ₂	Light	Photogating effect associated with trap states	Implementation of reconfigurable logic functions	N.A.	143
			Photoexcitation & charge trapping	Color-mixed pattern recognition	95%	144
				Collision detection	N.A.	126

context-aware experiences.¹²⁹ Advancing multimodal sensory computing requires a holistic approach inspired by biological sensory systems, such as the human retina, and the vision systems of flying animals and insects. This approach involves co-designing sensor technology, integration techniques, and computing architectures to optimize system performance, power consumption, and cost. Additionally, novel algorithms specifically designed for processing and interpreting multimodal sensory data are essential for enabling sophisticated perception, decision-making, and learning capabilities in these systems.

On-chip computing devices: conventional silicon CMOS electronics face limitations when implementing neural network algorithms due to their digital nature. Emerging neuromorphic computing devices such as two-terminal resistive switching memories are promising candidates for in-memory and in-sensor computing. Metal-oxide-semiconductor field effect transistors (MOSFETs) incorporating ferroelectric materials as dielectric components, known as negative capacitance field-effect transistors (NCFETs), offer potential advantages. Replacing the conventional insulated gate dielectric with a ferroelectric thin film exhibiting a “negative capacitance” effect can mitigate the short-channel effect that constrains integrated circuit performance. Ferroelectric materials are also well-suited for non-volatile memory applications due to their inherent spontaneous and remnant polarization properties.

Ferroelectric materials are commonly used in various sensing-memory-computing devices. For instance, in piezoelectric devices, all ferroelectric materials exhibit both ferroelectricity and piezoelectricity, facilitating the conversion between mechanical and electrical energy. Consequently, ferroelectric semiconductor materials can transduce information from both optical and mechanical signals. In available photodiodes, ferroelectric materials such as BiFeO_3 can be used to create optical devices with multifunctional capabilities. Despite significant advancements in applying ferroelectric materials for piezoelectric devices, research in other domains remains predominantly theoretical or experimental.

Low-temperature integration technology is essential for maintaining existing device functionalities during three-dimensional (3D) integration. Reliable low-temperature bonding and interconnect processes help minimize thermal expansion mismatches between stacked chips. Additionally, reducing the thickness of active devices and passive components is necessary to minimize parasitic time delays. Near/in-sensor architectures offer advantages in power efficiency, processing speed, and simplified circuitry. However, direct analog processing in these techniques is susceptible to noise, and the design of large-scale analog near-sensor systems is complex, limiting portability and scalability. Balancing noise, power, and design complexity is crucial for specific applications. Although there has been encouraging progress in emulating synapses and neurons using memristive devices, the field is still in its early stages. Experimental work is primarily at the single-device or small array level. Building practical hardware systems based on these prototypes requires significant scientific and engineering effort. Developing neuromorphic systems is highly

interdisciplinary, necessitating collaboration among physicists, material scientists, electrical engineers, and computer scientists. This collaboration is essential for understanding device mechanisms, designing material structures to control device behavior, and optimizing algorithms based on device characteristics.¹⁴⁶

5.2.2 3D integration of 2D electronics. In addition to the aforementioned challenges, advancing 3D integration of 2D electronics represents a critical pathway for the evolution of future 2D electronic systems. This approach involves the stacking of multiple 2D materials and devices to form a three-dimensional architecture, offering significant potential for both “More Moore” and “More than Moore” applications. The primary goals of 3D integration include increasing transistor density, advancing device scaling, and facilitating the development of multifunctional chips that combine logic, memory, sensing, and other functionalities within a single 3D integrated chip.¹⁴⁷ This strategy also aims to create sustainable and energy-efficient computing systems. Key methods of 3D Integration with 2D electronics include:

- **Stacked 2D FETs:** this method involves stacking multiple layers of 2D field-effect transistors (FETs) to optimize area utilization and achieve scaling benefits.
- **Monolithic 3D integration:** this technique involves integrating 2D materials directly with silicon-based logic or memory devices on a single chip, supporting the development of complex systems, such as image sensors and memory devices combined with traditional silicon circuits.

Significant advantages of 3D monolithic integration as a future technology. Increased interconnection density and bandwidth: 3D monolithic integration enables the fabrication of inter-layer vias with nanoscale diameters and pitches. This allows for a much higher interconnect density than through-silicon vias used in traditional 3D packaging techniques. This enhanced interconnect density enables high-bandwidth data transfer within the integrated circuit while minimizing power consumption, making it ideal for high-speed data processing and communication applications, such as artificial intelligence, high-performance computing, and edge computing.

Reduced electrical parasitic capacitance: the close proximity of components within a 3D monolithic integrated circuit shortens interconnect lengths, thereby reducing parasitic capacitance. This reduction in parasitic capacitance improves signal integrity and reduces signal delay, resulting in faster and more efficient circuits. Furthermore, the increased interconnect density, reduced parasitic capacitance, and the potential use of low-power materials like 2D semiconductors can significantly improve energy efficiency.¹⁴⁸

Optimal area utilization and enhanced scaling: 3D monolithic integration enables the vertical stacking of multiple device tiers, optimizing area usage to its fullest potential. This technique supports continued scaling beyond the constraints of conventional 2D planar technologies, facilitating the development of more intricate and potent devices without notably enlarging the footprint.

Integration of diverse functionalities: monolithic 3D integration serves as a platform for incorporating diverse functionalities

within a single chip, combining logic, memory, and sensing elements to create highly integrated systems with unique capabilities. This integration supports advanced applications, such as smart cameras with embedded artificial intelligence, intelligent robots for medical use, and edge computing devices with integrated processing and data storage. Consolidating multiple functionalities onto a single chip simplifies system architecture, reduces complexity, and improves overall performance.¹⁴⁹

Overall, 3D monolithic integration shows substantial promise for the future of electronics. Nonetheless, significant challenges remain, particularly in the application of 2D materials, which must be addressed for widespread adoption.

Advantages of 2DMs for 3D integration over silicon. Scalability and density: the ultrathin nature of 2D materials allows for denser vertical stacking than silicon, enabling higher device density and more efficient space utilization. Conventional silicon-based monolithic 3D integration, however, encounters challenges due to high processing temperatures (above 600 °C), which can damage lower tiers in a 3D stack. In contrast, 2D materials, particularly those that can be integrated at low temperatures, offer a solution to this limitation.

Performance and efficiency: the atomic thinness of 2D materials helps mitigate short channel effects, a key concern in scaled transistors, leading to improved performance and reduced leakage currents. Certain 2D materials, such as transition metal dichalcogenides (TMDs), even demonstrate high carrier mobilities, surpassing those of silicon, which enables faster switching speeds and enhances overall performance.

Back-end-of-line (BEOL) compatibility: the low-temperature processing requirements of many 2D materials make them compatible with BEOL integration within existing silicon CMOS processes. This compatibility allows 2D-based devices, such as sensors and memory, to be added onto prefabricated silicon circuits without risking damage to the underlying layers.¹⁵⁰

Multifunctional chips: the application of 2D materials in 3D integration facilitates the development of multifunctional chips that combine logic, memory, sensing, and additional functions within a single 3D-integrated framework.

Optoelectronics: materials such as graphene and molybdenum disulfide (MoS₂) have shown promise in applications like image sensors and photodetectors, highlighting their potential for advanced optoelectronic systems.

Memory devices: 2D materials, including MoS₂ and hexagonal boron nitride (h-BN), are promising candidates for non-volatile memory devices such as memristors and memtransistors, which can be seamlessly integrated with logic circuits.

Sensors: the high surface-to-volume ratio and unique properties of 2D materials render them particularly well-suited for sensing applications, including gas and biosensing. Integrating these sensors within 3D systems can enable the development of novel functionalities and applications.

5.2.3 Challenges of 2DMs in large-scale production and monolithic 3D integration. While 3D integration of 2D electronics holds transformative potential for enhancing device density, functionality, and energy efficiency, fully realizing these

benefits demands addressing significant challenges in the scalable production of 2DMs and establishing reliable monolithic 3D integration. This section outlines key challenges associated with using 2D materials for monolithic 3D integration.

Material synthesis and transfer: the direct growth of 2D materials on substrates at temperatures compatible with back-end-of-line (BEOL) processing (below 450 °C) presents a significant challenge. High temperatures are required for the high-quality growth of many 2D materials, particularly transition-metal dichalcogenides (TMDs), which risk damaging lower layers in a 3D stack. Preserving the structural integrity of large-area 2D films during transfer, while minimizing defects and contamination, is also essential. However, the atomically thin and fragile nature of 2D materials makes them susceptible to damage during transfer, leading to wrinkles, folds, cracks, and contamination, all of which can significantly degrade device performance and reliability.

Device fabrication and performance: achieving low-resistance ohmic contacts in 2D FETs remains a major obstacle, impacting device performance. The atomically thin nature of 2D materials and phenomena such as Fermi-level pinning complicate contact engineering. While progress have been made with metals like bismuth (Bi) and antimony (Sb),¹⁵¹ further research is required to develop reliable, low-resistance contact strategies that are compatible with large-scale fabrication and 3D integration.¹⁵²

High- κ dielectric integration: integrating high- κ dielectrics with 2D materials without introducing damage is crucial for achieving optimal performance in scaled devices, particularly in gate-all-around (GAA) architectures. The chosen dielectrics must have low leakage currents and high dielectric strength, and the integration process must minimize damage to the 2D material. Identifying suitable dielectric materials and developing damage-free integration processes are key areas of ongoing research.¹⁵²

Threshold voltage control: precise control over the threshold voltage in both n-type and p-type 2D FETs is essential for designing low-power, high-performance 3D CMOS circuits. Techniques such as substitutional doping and surface charge-transfer doping are being explored. However, achieving uniform doping density and precise control at large scales compatible with CMOS processes remains challenging.

Performance variability and yield: minimizing performance variations across individual devices and achieving high yields are crucial for reliable 3D integrated circuits. Due to the sensitivity of 2D materials to defects and process variations, optimizing synthesis, transfer, and fabrication methods to reduce defects and enhance uniformity is essential for commercial viability.

Heat dissipation bottleneck: the increased device density and power consumption in 3D integration generate considerable heat, necessitating efficient dissipation from upper tiers in a 3D stack to prevent performance degradation and ensure device reliability. However, the high out-of-plane thermal insulation of 2D materials, combined with challenges in integrating effective heat spreaders, complicates thermal management.

Thermal conductivity mismatch: the mismatch in thermal expansion coefficients between 2D materials and substrates

leads to stress and strain during temperature changes, affecting device reliability. Careful selection of materials and integration strategies is essential to mitigate these effects.

3D architecture and interconnects: creating reliable and low-resistance Inter-Tier Vias (ITVs) that connect different tiers in a 3D stack is essential for signal integrity and overall performance. Traditional high-temperature or high-energy ITV fabrication processes may damage 2D materials. Thus, developing low-temperature, damage-free processes compatible with 2D materials is critical.

Electrostatic coupling: the close proximity of conductive elements in densely packed 3D structures can lead to significant parasitic capacitance and electrostatic coupling, causing signal degradation, noise, and crosstalk. The atomically thin nature and high integration density achievable with 2D materials exacerbate these challenges. Mitigating these effects requires careful layout design, use of low- κ dielectrics, shielding layers, and proper grounding strategies.

Addressing these challenges is essential to unlock the full potential of 2D materials in monolithic 3D integration. The field is advancing rapidly, with ongoing research focused on developing new materials, fabrication processes, and integration strategies to overcome these hurdles and pave the way for commercially viable 2D-based 3D integrated circuits.

Author contributions

Y. S. and N. T. D. wrote the draft of the manuscript. Y. S., N. T. D., and K.-W. A. revised the manuscript. K.-W. A. conceived and supervised the project.

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by the National Research Foundation, Singapore, under the Competitive Research Program (NRF-CRP24-2020-0002, NRF-CRP22-2019-0007). N. T. D. acknowledges the BrainPool (BP) Program, National Research Foundation of Korea (NRF) (Grant Number: RS-2023-00264072).

References

- 1 T. P. Truong, H. T. Le and T. T. Nguyen, *J. Phys.: Conf. Ser.*, 2020, **1432**, 012068.
- 2 F. Zhou and Y. Chai, *Nat. Electron.*, 2020, **3**, 664–671.
- 3 Y. Chai, *Nature*, 2020, **579**, 32–33.
- 4 F. Liao, Z. Zhou, B. J. Kim, J. Chen, J. Wang, T. Wan, Y. Zhou, A. T. Hoang, C. Wang, J. Kang, J.-H. Ahn and Y. Chai, *Nat. Electron.*, 2022, **5**, 84–91.
- 5 N. T. Duong, Y. Shi, S. Li, Y.-C. Chien, H. Xiang, H. Zheng, P. Li, L. Li, Y. Wu and K.-W. Ang, *Adv. Sci.*, 2024, **11**, 2303447.
- 6 S. A. Han, T.-H. Kim, S. K. Kim, K. H. Lee, H.-J. Park, J.-H. Lee and S.-W. Kim, *Adv. Mater.*, 2018, **30**, 1800342.
- 7 Y. Wang, Y. Cai, F. Wang, J. Yang, T. Yan, S. Li, Z. Wu, X. Zhan, K. Xu, J. He and Z. Wang, *Nano Lett.*, 2023, **23**, 4524–4532.
- 8 Y. Wang, Y. Gong, S. Huang, X. Xing, Z. Lv, J. Wang, J.-Q. Yang, G. Zhang, Y. Zhou and S.-T. Han, *Nat. Commun.*, 2021, **12**, 5979.
- 9 P. P. Parlevliet, A. Kanaev, C. P. Hung, A. Schweiger, F. D. Gregory, R. Benosman, G. C. H. E. de Croon, Y. Gutfreund, C.-C. Lo and C. F. Moss, *Front. Neurosci.*, 2021, **15**, 672161.
- 10 X. Pang, Y. Wang, Y. Zhu, Z. Zhang, D. Xiang, X. Ge, H. Wu, Y. Jiang, Z. Liu, X. Liu, C. Liu, W. Hu and P. Zhou, *Nat. Commun.*, 2024, **15**, 1613.
- 11 Z. Zhang, S. Wang, C. Liu, R. Xie, W. Hu and P. Zhou, *Nat. Nanotechnol.*, 2022, **17**, 27–32.
- 12 L. Sun, S. Qu, Y. Du, L. Yang, Y. Li, Z. Wang and W. Xu, *ACS Photonics*, 2023, **10**, 242–252.
- 13 L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza and T. Mueller, *Nature*, 2020, **579**, 62–66.
- 14 Y. Zhong, J. Tang, X. Li, X. Liang, Z. Liu, Y. Li, Y. Xi, P. Yao, Z. Hao, B. Gao, H. Qian and H. Wu, *Nat. Electron.*, 2022, **5**, 672–681.
- 15 C.-Y. Wang, S.-J. Liang, S. Wang, P. Wang, Z. Li, Z. Wang, A. Gao, C. Pan, C. Liu, J. Liu, H. Yang, X. Liu, W. Song, C. Wang, B. Cheng, X. Wang, K. Chen, Z. Wang, K. Watanabe, T. Taniguchi, J. J. Yang and F. Miao, *Sci. Adv.*, 2020, **6**, eaba6173.
- 16 J.-K. Han, D.-M. Geum, M.-W. Lee, J.-M. Yu, S. K. Kim, S. Kim and Y.-K. Choi, *Nano Lett.*, 2020, **20**, 8781–8788.
- 17 D. Lee, M. Park, Y. Baek, B. Bae, J. Heo and K. Lee, *Nat. Commun.*, 2022, **13**, 5223.
- 18 G. Zhou, J. Li, Q. Song, L. Wang, Z. Ren, B. Sun, X. Hu, W. Wang, G. Xu, X. Chen, L. Cheng, F. Zhou and S. Duan, *Nat. Commun.*, 2023, **14**, 8489.
- 19 Y. Liu, Z. Ji, G. Cen, H. Sun, H. Wang, C. Zhao, Z. L. Wang and W. Mai, *Light: Sci. Appl.*, 2023, **12**, 43.
- 20 Q. Chen, Y. Zhang, S. Liu, T. Han, X. Chen, Y. Xu, Z. Meng, G. Zhang, X. Zheng, J. Zhao, G. Cao and G. Liu, *Adv. Intell. Syst.*, 2020, **2**, 2000122.
- 21 V. K. Dat, C. Hong, M. D. Tran, T. K. Chau, V. D. Do, T. T. Tran, M. C. Nguyen, H. P. Duong, S. Oh, W. J. Yu, J. Kim and J.-H. Kim, *Adv. Electron. Mater.*, 2024, **10**, 2300517.
- 22 L. Pi, P. Wang, S.-J. Liang, P. Luo, H. Wang, D. Li, Z. Li, P. Chen, X. Zhou, F. Miao and T. Zhai, *Nat. Electron.*, 2022, **5**, 248–254.
- 23 S. Seo, S.-H. Jo, S. Kim, J. Shim, S. Oh, J.-H. Kim, K. Heo, J.-W. Choi, C. Choi, S. Oh, D. Kuzum, H.-S. P. Wong and J.-H. Park, *Nat. Commun.*, 2018, **9**, 5106.

24 Y. Wu, N. T. Duong, Y.-C. Chien, S. Liu and K.-W. Ang, *Adv. Electron. Mater.*, 2024, **10**, 2300481.

25 F. Sun, H. Xu, W. Hong, Z. Sun and W. Liu, *Adv. Funct. Mater.*, 2024, **34**, 2313776.

26 N. Higashitarumizu, H. Kawamoto, C.-J. Lee, B.-H. Lin, F.-H. Chu, I. Yonemori, T. Nishimura, K. Wakabayashi, W.-H. Chang and K. Nagashio, *Nat. Commun.*, 2020, **11**, 2428.

27 Y. Li and J. Ibanez-Guzman, *IEEE Signal Process. Mag.*, 2020, **37**, 50–61.

28 S. Takemura, A. Bharioke, Z. Lu, A. Nern, S. Vitaladevuni, P. K. Rivlin, W. T. Katz, D. J. Olbris, S. M. Plaza, P. Winston, T. Zhao, J. A. Horne, R. D. Fetter, S. Takemura, K. Blazek, L.-A. Chang, O. Ogundeyi, M. A. Saunders, V. Shapiro, C. Sigmund, G. M. Rubin, L. K. Scheffer, I. A. Meinertzhagen and D. B. Chklovskii, *Nature*, 2013, **500**, 175–181.

29 B. Chakrabarti, M. A. Lastras-Montaño, G. Adam, M. Prezioso, B. Hoskins, M. Payvand, A. Madhavan, A. Ghofrani, L. Theogarajan, K.-T. Cheng and D. B. Strukov, *Sci. Rep.*, 2017, **7**, 42429.

30 C. Li, D. Belkin, Y. Li, P. Yan, M. Hu, N. Ge, H. Jiang, E. Montgomery, P. Lin, Z. Wang, W. Song, J. P. Strachan, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang and Q. Xia, *Nat. Commun.*, 2018, **9**, 2385.

31 F. Solis, Á. Fernández Bocco, A. C. Galetto, L. Passetti, M. R. Hueda and B. T. Reyes, *Int. J. Circuit Theory Appl.*, 2021, **49**, 3171–3185.

32 M. Bear, B. Connors and M. A. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain, Enhanced Edition*, Jones & Bartlett Learning, 2020.

33 T. Gollisch and M. Meister, *Neuron*, 2010, **65**, 150–164.

34 C. A. Curcio, K. R. Sloan, R. E. Kalina and A. E. Hendrickson, *J. Comp. Neurol.*, 1990, **292**, 497–523.

35 A. J. Oxenham, *Annu. Rev. Psychol.*, 2018, **69**, 27–50.

36 S. Ghosh, A. Pannone, D. Sen, A. Wali, H. Ravichandran and S. Das, *Nat. Commun.*, 2023, **14**, 6021.

37 H.-J. Tseng, W.-C. Tian and W.-J. Wu, *Sensors*, 2013, **13**, 5478–5492.

38 S. Gupta, L. Lorenzelli and R. Dahiya, in 2017 IEEE SENSORS, IEEE, Glasgow, 2017, pp. 1–3.

39 Y. Zhang, G. Lu, M. Chen, Y. Liu and R. Yao, *IEEE Sens. J.*, 2022, **22**, 12613–12621.

40 C. Liu, L. Wang, J. Qi and K. Liu, *Adv. Mater.*, 2020, **32**, 2000046.

41 X. Wu, R. Ge, P.-A. Chen, H. Chou, Z. Zhang, Y. Zhang, S. Banerjee, M.-H. Chiang, J. C. Lee and D. Akinwande, *Adv. Mater.*, 2019, **31**, 1806790.

42 X. Huang, C. Liu and P. Zhou, *npj 2D Mater. Appl.*, 2022, **6**, 51.

43 C. Liu, H. Chen, S. Wang, Q. Liu, Y.-G. Jiang, D. W. Zhang, M. Liu and P. Zhou, *Nat. Nanotechnol.*, 2020, **15**, 545–557.

44 D. Akinwande, C. Huyghebaert, C.-H. Wang, M. I. Serna, S. Goossens, L.-J. Li, H.-S. P. Wong and F. H. L. Koppens, *Nature*, 2019, **573**, 507–518.

45 Y. S. Ang, L. Cao and L. K. Ang, *InfoMat*, 2021, **3**, 502–535.

46 Z.-D. Luo, X. Xia, M.-M. Yang, N. R. Wilson, A. Gruverman and M. Alexe, *ACS Nano*, 2020, **14**, 746–754.

47 X. Wang, P. Wang, J. Wang, W. Hu, X. Zhou, N. Guo, H. Huang, S. Sun, H. Shen, T. Lin, M. Tang, L. Liao, A. Jiang, J. Sun, X. Meng, X. Chen, W. Lu and J. Chu, *Adv. Mater.*, 2015, **27**, 6538.

48 M.-Y. Tsai, C.-T. Huang, C.-Y. Lin, M.-P. Lee, F.-S. Yang, M. Li, Y.-M. Chang, K. Watanabe, T. Taniguchi, C.-H. Ho, W.-W. Wu, M. Yamamoto, J.-L. Wu, P.-W. Chiu and Y.-F. Lin, *Nat. Electron.*, 2023, **6**, 755–764.

49 Y. Chen, X. Wang, L. Huang, X. Wang, W. Jiang, Z. Wang, P. Wang, B. Wu, T. Lin, H. Shen, Z. Wei, W. Hu, X. Meng, J. Chu and J. Wang, *Nat. Commun.*, 2021, **12**, 4030.

50 Y. Xu, X. Xu, Y. Huang, Y. Tian, M. Cheng, J. Deng, Y. Xie, Y. Zhang, P. Zhang, X. Wang, Z. Wang, M. Li, L. Li and M. Liu, *Adv. Mater.*, 2024, **36**, 2402903.

51 J. Wang, Y. Fang, Y. Jiang, S. Long, X. Ma, X. Fu, Y. She, Q. Zhao, Y. Chen, H. Xu, T. Li, F. Zhong, H. Lin, X. Chen and J. Miao, *Appl. Phys. Lett.*, 2024, **124**, 121108.

52 W. Huh, S. Jang, J. Y. Lee, D. Lee, D. Lee, J. M. Lee, H.-G. Park, J. C. Kim, H. Y. Jeong, G. Wang and C.-H. Lee, *Adv. Mater.*, 2018, **30**, 1801447.

53 L. Sun, Y. Zhang, G. Han, G. Hwang, J. Jiang, B. Joo, K. Watanabe, T. Taniguchi, Y.-M. Kim, W. J. Yu, B.-S. Kong, R. Zhao and H. Yang, *Nat. Commun.*, 2019, **10**, 3161.

54 M. Wang, S. Cai, C. Pan, C. Wang, X. Lian, Y. Zhuo, K. Xu, T. Cao, X. Pan, B. Wang, S.-J. Liang, J. J. Yang, P. Wang and F. Miao, *Nat. Electron.*, 2018, **1**, 130–136.

55 J. H. Kim, S. Sarkar, Y. Wang, T. Taniguchi, K. Watanabe and M. Chhowalla, *Nano Lett.*, 2024, **24**, 2561–2566.

56 K. S. Novoselov, A. Mishchenko, A. Carvalho and A. H. Castro Neto, *Science*, 2016, **353**, aac9439.

57 M. Naqi, M. S. Kang, N. Liu, T. Kim, S. Baek, A. Bala, C. Moon, J. Park and S. Kim, *npj 2D Mater. Appl.*, 2022, **6**, 1–9.

58 J. Xie, S. Afshari and I. Sanchez Esqueda, *npj 2D Mater. Appl.*, 2022, **6**, 1–7.

59 S. Chen, M. R. Mahmoodi, Y. Shi, C. Mahata, B. Yuan, X. Liang, C. Wen, F. Hui, D. Akinwande, D. B. Strukov and M. Lanza, *Nat. Electron.*, 2020, **3**, 638–645.

60 S. Li, M.-E. Pam, Y. Li, L. Chen, Y.-C. Chien, X. Fong, D. Chi and K.-W. Ang, *Adv. Mater.*, 2022, **34**, 2103376.

61 T. Li, J. Miao, X. Fu, B. Song, B. Cai, X. Ge, X. Zhou, P. Zhou, X. Wang, D. Jariwala and W. Hu, *Nat. Nanotechnol.*, 2023, **18**, 1303–1310.

62 Y. Chen, X. Wang, G. Wu, Z. Wang, H. Fang, T. Lin, S. Sun, H. Shen, W. Hu, J. Wang, J. Sun, X. Meng and J. Chu, *Small*, 2018, **14**, 1703293.

63 Y. Yang, C. Pan, Y. Li, X. Yangdong, P. Wang, Z.-A. Li, S. Wang, W. Yu, G. Liu, B. Cheng, Z. Di, S.-J. Liang and F. Miao, *Nat. Electron.*, 2024, **7**, 225–233.

64 F. Liu, X. Lin, Y. Yan, X. Gan, Y. Cheng and X. Luo, *Nano Lett.*, 2023, **23**, 11645–11654.

65 D. Li, M. Chen, Z. Sun, P. Yu, Z. Liu, P. M. Ajayan and Z. Zhang, *Nat. Nanotechnol.*, 2017, **12**, 901–906.

66 M. Jiang, T. Zheng, J. Zhang, Z. Li, Q. Deng, Y. Pan, W. Gao, Y. Yang, J. Li and N. Huo, *Adv. Opt. Mater.*, 2024, **12**, 2303217.

67 J. Xiong, Q. Yu, X. Hou, B. Liu, S. Li, H. Deng, Z. Yang, J. Leng, S. Zhu, Y. Sun, Z. Jiang, N. Huo, J. Wu and P. Zhou, *Adv. Funct. Mater.*, 2024, **34**, 2314972.

68 J. Lee, N. T. Duong, S. Bang, C. Park, D. A. Nguyen, H. Jeon, J. Jang, H. M. Oh and M. S. Jeong, *Nano Lett.*, 2020, **20**, 2370–2377.

69 C. Choi, J. Leem, M. Kim, A. Taqieddin, C. Cho, K. W. Cho, G. J. Lee, H. Seung, H. J. Bae, Y. M. Song, T. Hyeon, N. R. Aluru, S. Nam and D.-H. Kim, *Nat. Commun.*, 2020, **11**, 5934.

70 G. Feng, Q. Zhu, X. Liu, L. Chen, X. Zhao, J. Liu, S. Xiong, K. Shan, Z. Yang, Q. Bao, F. Yue, H. Peng, R. Huang, X. Tang, J. Jiang, W. Tang, X. Guo, J. Wang, A. Jiang, B. Dkhil, B. Tian, J. Chu and C. Duan, *Nat. Commun.*, 2024, **15**, 513.

71 H. Ning, Z. Yu, Q. Zhang, H. Wen, B. Gao, Y. Mao, Y. Li, Y. Zhou, Y. Zhou, J. Chen, L. Liu, W. Wang, T. Li, Y. Li, W. Meng, W. Li, Y. Li, H. Qiu, Y. Shi, Y. Chai, H. Wu and X. Wang, *Nat. Nanotechnol.*, 2023, **18**, 493–500.

72 Y.-C. Chien, H. Xiang, Y. Shi, N. T. Duong, S. Li and K.-W. Ang, *Adv. Mater.*, 2023, **35**, 2204949.

73 H. Xiang, Y. Chien, L. Li, H. Zheng, S. Li, N. T. Duong, Y. Shi and K. Ang, *Adv. Funct. Mater.*, 2023, **33**, 2304657.

74 L. Tong, Z. Peng, R. Lin, Z. Li, Y. Wang, X. Huang, K.-H. Xue, H. Xu, F. Liu, H. Xia, P. Wang, M. Xu, W. Xiong, W. Hu, J. Xu, X. Zhang, L. Ye and X. Miao, *Science*, 2021, **373**, 1353–1358.

75 T. Jin, J. Mao, J. Gao, C. Han, K. P. Loh, A. T. S. Wee and W. Chen, *ACS Nano*, 2022, **16**, 13595–13611.

76 R. Hao, X. Hu, L. Luo, X. Yang, Q. Feng, Y. Li and Z. Zhang, *IEEE Electron Device Lett.*, 2024, **45**, 1080–1083.

77 Y. Shi, N. T. Duong, Y.-C. Chien, S. Li, H. Xiang, H. Zheng and K.-W. Ang, *Adv. Funct. Mater.*, 2024, **34**, 2400879.

78 C. Chen, Y. Zhou, L. Tong, Y. Pang and J. Xu, *Adv. Mater.*, 2024, 2400332.

79 J. Junquera and P. Ghosez, *Nature*, 2003, **422**, 506–509.

80 J. Gao, X. Lian, Z. Chen, S. Shi, E. Li, Y. Wang, T. Jin, H. Chen, L. Liu, J. Chen, Y. Zhu and W. Chen, *Adv. Funct. Mater.*, 2022, **32**, 2110415.

81 L. Jin, H. Wang, R. Cao, K. Khan, A. K. Tareen, S. Wageh, A. A. Al-Ghamdi, S. Li, D. Li, Y. Zhang and H. Zhang, *APL Mater.*, 2022, **10**, 060903.

82 D. G. Schlom, L.-Q. Chen, C.-B. Eom, K. M. Rabe, S. K. Streiffer and J.-M. Triscone, *Annu. Rev. Mater. Res.*, 2007, **37**, 589–626.

83 M. H. Yusuf, B. Nielsen, M. Dawber and X. Du, *Nano Lett.*, 2014, **14**, 5437–5444.

84 F. Yang, H. K. Ng, X. Ju, W. Cai, J. Cao, D. Chi, A. Suwardi, G. Hu, Z. Ni, X. R. Wang, J. Lu and J. Wu, *Adv. Funct. Mater.*, 2024, **34**, 2310438.

85 F. Xue, J.-H. He and X. Zhang, *Appl. Phys. Rev.*, 2021, **8**, 021316.

86 D. Zhang, Z.-D. Luo, Y. Yao, P. Schoenherr, C. Sha, Y. Pan, P. Sharma, M. Alexe and J. Seidel, *Nano Lett.*, 2021, **21**, 995–1002.

87 J. Chen, C. Zhu, G. Cao, H. Liu, R. Bian, J. Wang, C. Li, J. Chen, Q. Fu, Q. Liu, P. Meng, W. Li, F. Liu and Z. Liu, *Adv. Mater.*, 2022, **34**, 2104676.

88 F. Liu, L. You, K. L. Seyler, X. Li, P. Yu, J. Lin, X. Wang, J. Zhou, H. Wang, H. He, S. T. Pantelides, W. Zhou, P. Sharma, X. Xu, P. M. Ajayan, J. Wang and Z. Liu, *Nat. Commun.*, 2016, **7**, 12357.

89 P. Wang, J. Li, W. Xue, W. Ci, F. Jiang, L. Shi, F. Zhou, P. Zhou and X. Xu, *Adv. Sci.*, 2024, **11**, 2305679.

90 J. Zeng, G. Feng, G. Wu, J. Liu, Q. Zhao, H. Wang, S. Wu, X. Wang, Y. Chen, S. Han, B. Tian, C. Duan, T. Lin, J. Ge, H. Shen, X. Meng, J. Chu and J. Wang, *Adv. Funct. Mater.*, 2024, **34**, 2313010.

91 B. Das, S. Baek, J. Niu, C. Jang, Y. Lee and S. Lee, *ACS Nano*, 2023, **17**, 21297–21306.

92 S. Wang, L. Liu, L. Gan, H. Chen, X. Hou, Y. Ding, S. Ma, D. W. Zhang and P. Zhou, *Nat. Commun.*, 2021, **12**, 53.

93 M. Si, A. K. Saha, S. Gao, G. Qiu, J. Qin, Y. Duan, J. Jian, C. Niu, H. Wang, W. Wu, S. K. Gupta and P. D. Ye, *Nat. Electron.*, 2019, **2**, 580–586.

94 C. Cui, W.-J. Hu, X. Yan, C. Addiego, W. Gao, Y. Wang, Z. Wang, L. Li, Y. Cheng, P. Li, X. Zhang, H. N. Alshareef, T. Wu, W. Zhu, X. Pan and L.-J. Li, *Nano Lett.*, 2018, **18**, 1253–1258.

95 W. Ding, J. Zhu, Z. Wang, Y. Gao, D. Xiao, Y. Gu, Z. Zhang and W. Zhu, *Nat. Commun.*, 2017, **8**, 14956.

96 J. Xiao, H. Zhu, Y. Wang, W. Feng, Y. Hu, A. Dasgupta, Y. Han, Y. Wang, D. A. Muller, L. W. Martin, P. Hu and X. Zhang, *Phys. Rev. Lett.*, 2018, **120**, 227601.

97 M. Gabel and Y. Gu, *Adv. Funct. Mater.*, 2021, **31**, 2009999.

98 Y. Wang, Y. Gong, L. Yang, Z. Xiong, Z. Lv, X. Xing, Y. Zhou, B. Zhang, C. Su, Q. Liao and S.-T. Han, *Adv. Funct. Mater.*, 2021, **31**, 2100144.

99 L. Liu, Y. Zhang and Y. Yan, *Nanoscale Horiz.*, 2023, **8**, 1301–1312.

100 T. Wan, B. Shao, S. Ma, Y. Zhou, Q. Li and Y. Chai, *Adv. Mater.*, 2023, **35**, 2203830.

101 Z. Shen, C. Zhao, L. Kang, Y. Sun, Y. Liu, I. Z. Mitrovic, L. Yang, E. G. Lim, Z. Zeng and C. Zhao, *Adv. Intell. Syst.*, 2022, **4**, 2100236.

102 J. Wang, N. Ilyas, Y. Ren, Y. Ji, S. Li, C. Li, F. Liu, D. Gu and K. Ang, *Adv. Mater.*, 2024, **36**, 2307393.

103 J. Chen, Z. Zhou, B. J. Kim, Y. Zhou, Z. Wang, T. Wan, J. Yan, J. Kang, J.-H. Ahn and Y. Chai, *Nat. Nanotechnol.*, 2023, **18**, 882–888.

104 J. Liang, X. Yu, J. Qiu, M. Wang, C. Cheng, B. Huang, H. Zhang, R. Chen, W. Pei and H. Chen, *ACS Appl. Mater. Interfaces*, 2023, **15**, 9584–9592.

105 L. F. Abbott and W. G. Regehr, *Nature*, 2004, **431**, 796–803.

106 L. Hu, J. Yang, J. Wang, P. Cheng, L. O. Chua and F. Zhuge, *Adv. Funct. Mater.*, 2021, **31**, 2005582.

107 Y. Zhong, J. Tang, X. Li, B. Gao, H. Qian and H. Wu, *Nat. Commun.*, 2021, **12**, 408.

108 C. Gao, D. Liu, C. Xu, W. Xie, X. Zhang, J. Bai, Z. Lin, C. Zhang, Y. Hu, T. Guo and H. Chen, *Nat. Commun.*, 2024, **15**, 740.

109 M. Yan, C. Huang, P. Bienstman, P. Tino, W. Lin and J. Sun, *Nat. Commun.*, 2024, **15**, 2056.

110 M. Lukoševičius and H. Jaeger, *Comput. Sci. Rev.*, 2009, **3**, 127–149.

111 W. Maass, T. Natschläger and H. Markram, *Neural Comput.*, 2002, **14**, 2531–2560.

112 J. Cao, X. Zhang, H. Cheng, J. Qiu, X. Liu, M. Wang and Q. Liu, *Nanoscale*, 2022, **14**, 289–298.

113 L. Appeltant, M. C. Soriano, G. Van Der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso and I. Fischer, *Nat. Commun.*, 2011, **2**, 468.

114 C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee and W. D. Lu, *Nat. Commun.*, 2017, **8**, 2204.

115 X. Zhu, Q. Wang and W. D. Lu, *Nat. Commun.*, 2020, **11**, 2439.

116 K. Liu, T. Zhang, B. Dang, L. Bao, L. Xu, C. Cheng, Z. Yang, R. Huang and Y. Yang, *Nat. Electron.*, 2022, **5**, 761–773.

117 J. Zha, Y. Xia, S. Shi, H. Huang, S. Li, C. Qian, H. Wang, P. Yang, Z. Zhang, Y. Meng, W. Wang, Z. Yang, H. Yu, J. C. Ho, Z. Wang and C. Tan, *Adv. Mater.*, 2024, **36**, 2308502.

118 L. Sun, Z. Wang, J. Jiang, Y. Kim, B. Joo, S. Zheng, S. Lee, W. J. Yu, B.-S. Kong and H. Yang, *Sci. Adv.*, 2021, **7**, eabg1455.

119 G. Feng, X. Zhang, B. Tian and C. Duan, *InfoMat*, 2023, **5**, e12473.

120 P. Wu, T. He, H. Zhu, Y. Wang, Q. Li, Z. Wang, X. Fu, F. Wang, P. Wang, C. Shan, Z. Fan, L. Liao, P. Zhou and W. Hu, *InfoMat*, 2022, **4**, e12275.

121 S. Judge and F. Rind, *J. Exp. Biol.*, 1997, **200**, 2209–2216.

122 F. Rind, *J. Exp. Biol.*, 1984, **110**, 143–167.

123 R. R. Harrison, *IEEE Trans. Circuits Syst. I*, 2005, **52**, 2308–2318.

124 H. Okuno and T. Yagi, *Rob. Auton. Syst.*, 2009, **57**, 508–516.

125 R. Takami, K. Shimonomura, S. Kameda and T. Yagi, in *2005 IEEE International Symposium on Circuits and Systems*, IEEE, Kobe, Japan, 2005, pp. 2771–2774.

126 D. Jayachandran, A. Oberoi, A. Sebastian, T. H. Choudhury, B. Shankar, J. M. Redwing and S. Das, *Nat. Electron.*, 2020, **3**, 646–655.

127 H. Kim, S. Oh, H. Choo, D.-H. Kang and J.-H. Park, *ACS Nano*, 2023, **17**, 17332–17341.

128 Y. Zhang, J. Mao, R.-K. Zheng, J. Zhang, Y. Wu, X. Wang, K. Miao, H. Yao, L. Yang and H. Zheng, *Adv. Funct. Mater.*, 2023, **33**, 2214745.

129 L. Hu, B. J. Kim, S. Ji, J. Hong, A. K. Katiyar and J.-H. Ahn, *Appl. Phys. Rev.*, 2022, **9**, 041308.

130 J. Huang, J. Feng, Z. Chen, Z. Dai, S. Yang, Z. Chen, H. Zhang, Z. Zhou, Z. Zeng, X. Li and X. Gui, *Nano Energy*, 2024, **126**, 109684.

131 K. I. Bolotin, K. J. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim and H. L. Stormer, *Solid State Commun.*, 2008, **146**, 351–355.

132 D. Kuzum, H. Takano, E. Shim, J. C. Reed, H. Juul, A. G. Richardson, J. de Vries, H. Bink, M. A. Dichter, T. H. Lucas, D. A. Coulter, E. Cubukcu and B. Litt, *Nat. Commun.*, 2014, **5**, 5259.

133 N. Alzate-Carvajal and A. Luican-Mayer, *ACS Omega*, 2020, **5**, 21320–21329.

134 X. Li, S. Li, J. Tian, F. Lyu, J. Liao and Q. Chen, *Adv. Funct. Mater.*, 2024, **34**, 2306486.

135 G. Wu, X. Zhang, G. Feng, J. Wang, K. Zhou, J. Zeng, D. Dong, F. Zhu, C. Yang, X. Zhao, D. Gong, M. Zhang, B. Tian, C. Duan, Q. Liu, J. Wang, J. Chu and M. Liu, *Nat. Mater.*, 2023, **22**, 1499–1506.

136 F. Wang, F. Hu, M. Dai, S. Zhu, F. Sun, R. Duan, C. Wang, J. Han, W. Deng, W. Chen, M. Ye, S. Han, B. Qiang, Y. Jin, Y. Chua, N. Chi, S. Yu, D. Nam, S. H. Chae, Z. Liu and Q. J. Wang, *Nat. Commun.*, 2023, **14**, 1938.

137 N. Jiang, J. Tang, W. Zhang, Y. Li, N. Li, X. Li, X. Chen, R. Fang, Z. Guo, F. Wang, J. Wang, Z. Li, C. He, G. Zhang, Z. Wang and D. Shang, *Adv. Opt. Mater.*, 2023, **11**, 2300271.

138 Y. Wang, H. Sun, Z. Sheng, J. Dong, W. Hu, D. Tang and Z. Zhang, *Nano Res.*, 2023, **16**, 12713–12719.

139 Y. Zhou, J. Fu, Z. Chen, F. Zhuge, Y. Wang, J. Yan, S. Ma, L. Xu, H. Yuan, M. Chan, X. Miao, Y. He and Y. Chai, *Nat. Electron.*, 2023, **6**, 870–878.

140 Q. Yang, Z.-D. Luo, D. Zhang, M. Zhang, X. Gan, J. Seidel, Y. Liu, Y. Hao and G. Han, *Adv. Funct. Mater.*, 2022, **32**, 202207290.

141 Y. Gong, P. Xie, X. Xing, Z. Lv, T. Xie, S. Zhu, H.-H. Hsu, Y. Zhou and S.-T. Han, *Adv. Funct. Mater.*, 2023, **33**, 2303539.

142 S. Dhara, H. Jawa, S. Ghosh, A. Varghese, D. Karmakar and S. Lodha, *ACS Appl. Mater. Interfaces*, 2021, **13**, 30785–30796.

143 R. Kumar, P. K. Kulriya, M. Mishra, F. Singh, G. Gupta and M. Kumar, *Nanotechnology*, 2018, **29**, 464001.

144 T. Järvinen, G. S. Lorite, J. Peräntie, G. Toth, S. Saarakkala, V. K. Virtanen and K. Kordas, *Nanotechnology*, 2019, **30**, 405501.

145 Z. Wang, T. Wan, S. Ma and Y. Chai, *Nat. Nanotechnol.*, 2024, 1–12.

146 Q. Duan, Z. Xiao, K. Yang and Y. Yang, in *Near-sensor and In-sensor Computing*, ed. Y. Chai and F. Liao, Springer International Publishing, Cham, 2022, pp. 1–31.

147 Y. Guo, J. Li, X. Zhan, C. Wang, M. Li, B. Zhang, Z. Wang, Y. Liu, K. Yang, H. Wang, W. Li, P. Gu, Z. Luo, Y. Liu, P. Liu, B. Chen, K. Watanabe, T. Taniguchi, X.-Q. Chen, C. Qin, J. Chen, D. Sun, J. Zhang, R. Wang, J. Liu, Y. Ye, X. Li, Y. Hou, W. Zhou, H. Wang and Z. Han, *Nature*, 2024, **630**, 346–352.

148 D. Lu, Y. Chen, Z. Lu, L. Ma, Q. Tao, Z. Li, L. Kong, L. Liu, X. Yang, S. Ding, X. Liu, Y. Li, R. Wu, Y. Wang, Y. Hu, X. Duan, L. Liao and Y. Liu, *Nature*, 2024, **630**, 340–345.

149 S. Kim, J. Seo, J. Choi and H. Yoo, *Nano-Micro Lett.*, 2022, **14**, 201.

150 K.-H. Kim, S. Oh, M. M. A. Fiagbenu, J. Zheng, P. Musavigharavi, P. Kumar, N. Trainor, A. Aljarb, Y. Wan, H. M. Kim, K. Katti, S. Song, G. Kim, Z. Tang, J.-H. Fu, M. Hakami, V. Tung, J. M. Redwing, E. A. Stach, R. H. Olsson and D. Jariwala, *Nat. Nanotechnol.*, 2023, **18**, 1044–1050.

151 D. Jayachandran, R. Pendurthi, M. U. K. Sadaf, N. U. Sakib, A. Pannone, C. Chen, Y. Han, N. Trainor, S. Kumari, T. V. Mc Knight, J. M. Redwing, Y. Yang and S. Das, *Nature*, 2024, **625**, 276–281.

152 D. Jayachandran, N. U. Sakib and S. Das, *Nat. Rev. Electr. Eng.*, 2024, **1**, 300–316.