

In-Sensor Computing: Materials, Devices, and Integration Technologies

Tianqing Wan, Bangjie Shao, Sijie Ma, Yue Zhou, Qiao Li, and Yang Chai*

The number of sensor nodes in the Internet of Things is growing rapidly, leading to a large volume of data generated at sensory terminals. Frequent data transfer between the sensors and computing units causes severe limitations on the system performance in terms of energy efficiency, speed, and security. To efficiently process a substantial amount of sensory data, a novel computation paradigm that can integrate computing functions into sensor networks should be developed. The in-sensor computing paradigm reduces data transfer and also decreases the high computing complexity by processing data locally. Here, the hardware implementation of the in-sensor computing paradigm at the device and array levels is discussed. The physical mechanisms that lead to unique sensory response characteristics and their corresponding computing functions are illustrated. In particular, bioinspired device characteristics enable the implementation of the functionalities of neuromorphic computation. The integration technology is also discussed and the perspective on the future development of in-sensor computing is provided.

1. Introduction

The number of sensor nodes has increased rapidly in the past few years (40 billion by 2022) and this trend will continue in the future.^[1–3] In a conventional sensory system (Figure 1a), the sensors generate raw and unstructured data, which are then transferred to the subsequent computing units and memory. The physical separation between the sensing, computing, and memory units causes inevitable data conversion and transfer. The corresponding time latency, energy consumption, and security threats are becoming intolerable in this era of big data. To further increase the efficiency and performance of the sensory system, significant innovations in materials, device physics, and computing architectures are required.

As the size of Si-based transistors continues to scale down, more complex and low-power circuits can be fabricated near the


sensory terminals, which can preprocess the sensor outputs for more efficient data transfer and processing.^[4–8] Although this near-sensor computing paradigm significantly reduces redundant data, physical separation still exists between the sensors and computing units, deteriorating speed and energy efficiency.^[9–12] A more promising approach is to completely eliminate this physical separation by integrating computing functions into the sensor (Figure 1b). In this in-sensor computing paradigm, the sensors not only transduce external stimuli to electrical outputs but also efficiently process the stimuli.^[10,13–16] The fusion of sensing and computing functions in one physical element can significantly reduce data transfer and simplify the system structure.

In addition to the benefits of time latency and energy consumption, the in-sensor computing paradigm also has a high computing complexity because of the sensor array structure. Within the array, multiple sensor elements can simultaneously collect external stimuli.^[17–19] Owing to the inherent computing capability of sensors, multiple computations are executed simultaneously. This high parallelism significantly increases the computing throughput. In addition, physical coupling between the sensor elements enables more computational functions. For example, the currents from individual sensors are summed according to Kirchhoff's laws, which naturally realize matrix summation. This matrix summation plays an essential role in receptive fields and neural networks.

At both the single-device and array levels, the in-sensor computing paradigm can provide numerous benefits to system performance. To explore the inherent computing capability of the sensor to reduce data transfer, a thorough investigation of the interaction between external stimuli and materials is required, as well as new insights into the physics, materials, device structure, and integration technologies. Accordingly, we discuss the in-sensor computing paradigm at the device and array levels because these application scenarios involve different considerations in materials, devices, and system structures. Specifically, we explore the physical mechanism that leads to unique response characteristics and corresponding computing functions, such as feature enhancement, nociceptors, highly secured cryptography, edge extraction, and pattern recognition. Finally, we discuss important points related to system integration and provide our perspective on future directions for in-sensor computing.

T. Wan, B. Shao, S. Ma, Y. Zhou, Q. Li, Y. Chai
Department of Applied Physics
The Hong Kong Polytechnic University
Hong Kong, China
E-mail: ychai@polyu.edu.hk

Y. Chai
Shenzhen Research Institute
The Hong Kong Polytechnic University
Shenzhen, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.202203830>.

DOI: 10.1002/adma.202203830

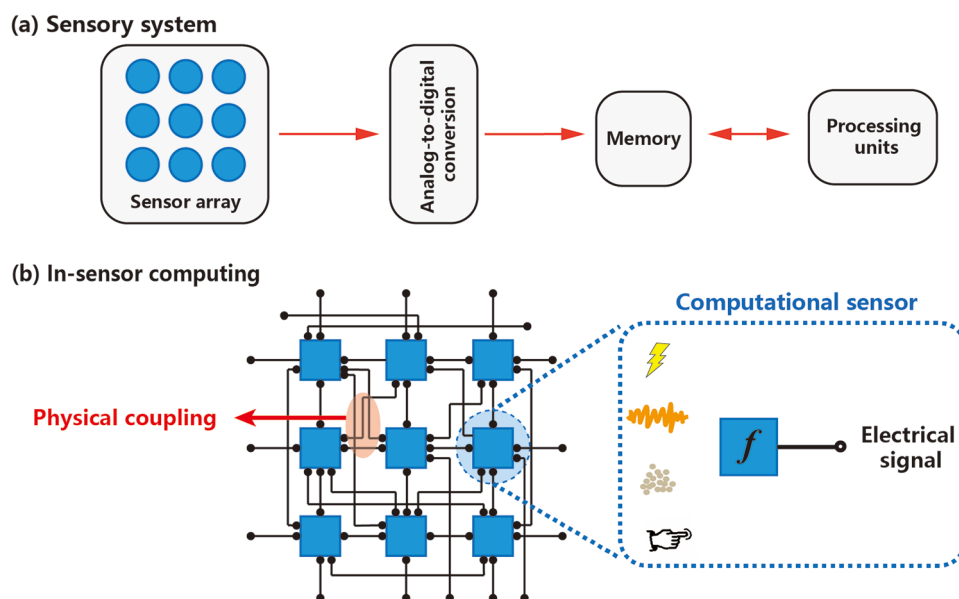


Figure 1. a) The architecture of conventional sensory system. The analog outputs from the sensory array are first converted to digital signals, then transferred to memory and computing units for processing. b) The schematic of the sensor array for in-sensor computing. At the device level, the computational sensors can execute computing functions during the transduction of external stimuli. At the array level, the physical coupling between sensor elements contributes to higher computing complexity.

2. In-Sensor Computing at the Device Level

Conventional sensors output real-time electrical signals proportional to the intensity of the input stimuli, whereas back-end computation circuits extract important temporal and spatial information. In the early ages of sensory computing, it was efficient to divide the computation tasks into separated hardware because the amount of sensory data was relatively limited. In addition, sensors and back-end computation circuits typically involve different materials, device structures, and manufacturing technologies. Their integration with conventional packaging technology has accelerated the design and implementation of sensory computing systems. In the era of big data, the input of substantial raw and unstructured data from sensors to back-end circuits results in unavoidable analog-to-digital signal conversion and data transfer, which causes a heavy burden on the transmission bandwidth and power consumption.

In contrast, sensors with an inherent computing ability can execute signal conversion and information processing at the same physical unit, which significantly reduces data transfer and simplifies the system structure. In this computational sensor, the magnitude of the electrical outputs is not linearly proportional to the input stimulus intensity, meaning the sensor outputs do not longer faithfully reproduce the original stimuli. They become more concise representations that contain important temporal, spatial, or relationship information. Specific computing functions are derived directly from the relationships between stimuli and the output magnitude, which are determined by the response characteristics of the materials. The underlying physical processes behind the interaction between external stimuli and materials should be thoroughly investigated to realize the desired response characteristics. Here, we first introduce computing functions and then discuss their direct realization through the response characteristics of sensory devices.

2.1. Feature Enhancement

The most basic and frequent task for the sensory system is to recognize patterns from complex environments, such as objects, sounds, and odors. The noisy environments render the accurate extraction of patterns difficult and costly. Researchers have adopted various sophisticated circuits for feature enhancement and pattern recognition to increase system performance. Feature enhancement functions are usually placed close to the sensor terminals. As the patterns become more distinguishable, the workload of the subsequent processing units for accurate pattern recognition is reduced, thereby reducing the overall time and energy consumption. The simplest strategy for feature enhancement is to increase the responsivity over the entire intensity range, which amplifies the output differences between different intensity signals. However, this simple strategy exhibits poor performance in terms of efficiency and signal-to-noise ratio, which requires more complex processing.

2.1.1. Superlinear Relationship

A more efficient strategy is to process signals from noisy backgrounds and objects selectively. Particularly, low-intensity noise signals are suppressed, and high-intensity feature signals are enhanced. The output from the feature-enhancement circuits happens to exhibit a superlinear relationship with the stimulus intensity (the red line in **Figure 2a**, $\frac{d \ln I}{d \ln P} > 1$, I represents the output magnitude and P represents the stimulus intensity). The slope increases with an increase in the stimulation intensity. In the near-sensor computing paradigm, the sensors and feature enhancement circuits are separated in physical space,

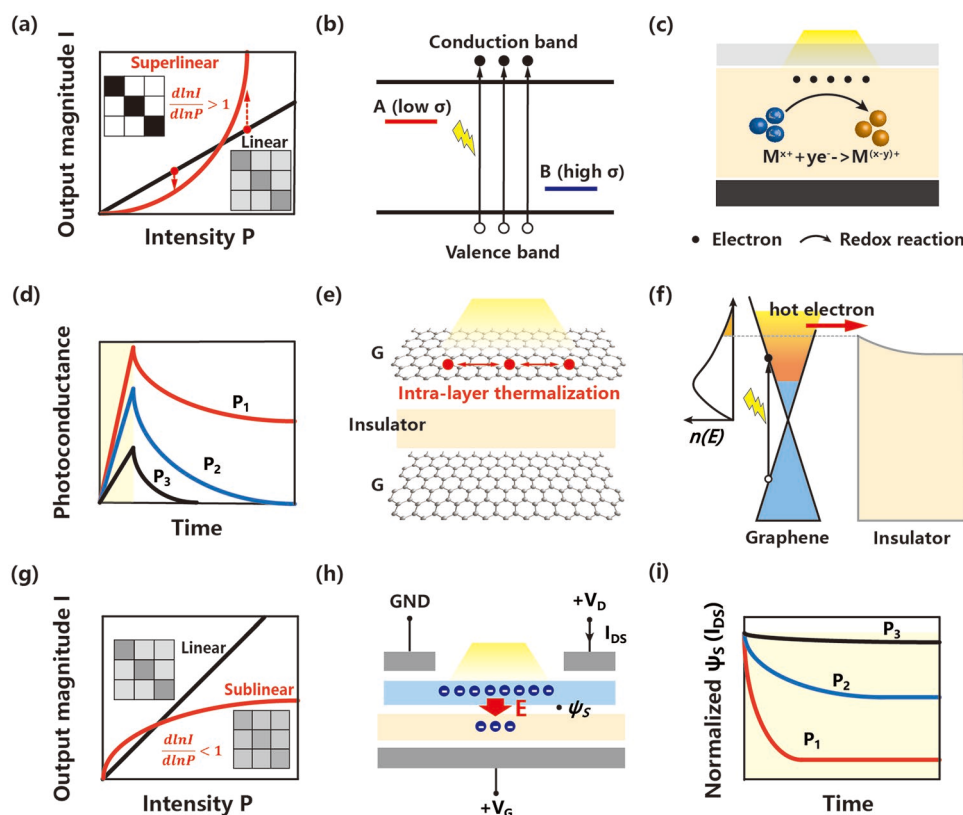


Figure 2. a) Feature enhancement through superlinear relationship between stimulus intensity and output magnitude. Compared to linear relationship ($\frac{d\ln I}{d\ln P} = 1$), the output magnitude from low (high)-intensity stimuli is suppressed (enhanced). b) Schematic illustration of recombination centers and their capture probability σ . c) Illustration of the device structure and electrochemical reactions. The light illumination generates substantial electrons, which facilitate the redox reactions. d) As evolved with time, the low photoconductance relaxes quickly, whereas the high photoconductance still remains at a high level. e) Illustration of the device structure for photoassisted thermionic carrier emission. f) Illustration of the energy band structure at the interface. A high barrier between graphene and insulator determines the carrier transport process. The photoexcited electrons thermalize quickly to have a much higher electron temperature than the lattice, which raises the electron distribution. More hot electrons can overcome the energy barrier. g) The illustration of linear and sub-linear relationship between intensity and output magnitude. Under sublinear relationship, the features become blurrier. h) The device structure of n-type MOSFET. Under positive gate bias ($+V_G$), the electrons in channel can overcome the interfacial barrier and get trapped in the gate dielectrics. i) Under light illumination, the surface potential on the conductive channel (ψ_s) decreases with time. With higher intensity, the degree of potential decrease is higher. The channel current (I_{DS}) shares similar trends with the surface potential ψ_s .

leading to inevitable data conversion and transfer. This physical separation can be completely alleviated by endowing the sensor with an in situ feature enhancement function, which requires a direct superlinear relationship between the stimulus intensity and sensor output. However, common sensors (such as light, pressure, temperature, and magnetic sensors) usually exhibit linear or even sublinear response characteristics. The realization of superlinear response characteristics is relatively challenging, requiring careful investigation of device physics and materials. The detailed material composition and device structure may vary for the different types of sensors. However, to realize feature enhancement, the superlinear relationship between stimulus intensity and output magnitude applies to all stimuli.

Because theoretical models and quantitative calculations have been established for various photoelectric effects, we illustrate the photoresponse characteristics to clarify the superlinear relationship for computational sensors. Light illumination can generate free electrons and holes in semiconductors, leading

to an increase in conductance. The free carrier concentration results from the dynamic equilibrium between excitation and recombination. At low light intensities, the small number of non-equilibrium carriers has a negligible impact on the recombination probability (σ), leading to a linear relationship between the photocurrent and light intensity. At very high light intensities, the recombination probability increases significantly, leading to a slower increase in carrier concentration. To facilitate the increase in carrier concentration and realize superlinear photoconductivity, the recombination probability should decrease with an increase in light intensity. This phenomenon requires at least two competing recombination centers with different capture probabilities (Figure 2b).^[20–24] In a specific intensity range, the recombination traffic shifts from centers with a high capture probability to centers with a low capture probability. With an increased light intensity, the recombination process is suppressed, which requires a higher concentration of free carriers to achieve an equilibrium between excitation and recombination. Therefore, the carrier concentration exhibits

a superlinear relationship with the light intensity. Quantitatively, the intensity range for superlinear photoconductivity ($d\ln I/d\ln P > 1$) and the formulation of photoconductivity with light intensity can be calculated, which varies depending on the temperature, capture probability, and distribution of energy levels. Through careful defect engineering, superlinear photoconductivity can be realized in a modest light-intensity range. Compared with the original stimuli pattern, the output magnitude from the high-intensity stimuli is significantly enhanced. Therefore, the photoconductors in situ realize feature enhancement functions during light-electrical signal conversion.

In the aforementioned scenario, the photoexcited carriers are trapped by defects in the materials and retained over a long time before recombination. Therefore, their contribution to the photoconductance is maintained for a longer time. Following the same strategy, the photoexcited carriers can facilitate photoconductance by involving electrochemical reactions, leading to a change in the valence state or lattice structure of the photoactive materials, as shown in Figure 2c.^[14,25–28] The corresponding electronic properties, such as conductivity, also change and are maintained for a relatively long time, which forms the physical basis of optoelectronic resistive switching memory. At higher light intensities, photoconductors usually switch to a more conductive state, which is maintained over a longer time. The combination of a higher photoconductance and retention can also lead to a superlinear relationship between conductance and light intensity. We used MoO₃ as the photoactive material, the conductivity of which depended on the Mo valence state.^[14] Under light illumination, the photoexcited carriers reacted with moisture and generate H⁺, which reduced MoO_x to a more conductive state (Mo⁶⁺ to Mo⁵⁺). The resultant H_yMoO_x lattice was relatively stable and was maintained for a long time after the removal of illumination. The retention increased with light intensity, meaning the rate of conductance decrease was lower. As evolved with time, the low conductance state relaxed quickly, whereas the high conductance state remained at a high level, as shown in Figure 2d. Therefore, the resultant conductance exhibited a superlinear relationship with the light intensity. We could realize feature enhancement functions inside the sensor array.

The second strategy to realize superlinear photoconductivity is thermionic carrier emission across the interfacial barrier.^[29,30] As the electrical bias can modulate the height of the interfacial barrier, the current exhibits a superlinear relationship with the electrical bias. In contrast to energy-band modulation, photoassisted thermionic carrier emission changes the carrier distribution by modulating the carrier temperature. A high electron temperature can be realized through light stimulation in graphene with strong electron-electron interactions. As shown in Figure 2e, the device structure comprises two graphene layers sandwiching a tunneling layer (such as h-BN, WSe₂), in which graphene functions as the photoactive layer and an interfacial barrier determines the transport properties. The photoexcited electrons thermalize quickly to have a much higher electron temperature than that of the lattice. As the elevated temperature raises the electron distribution, more electrons can overcome the interfacial barrier, resulting in a superlinear relationship between the photocurrent and the light intensity (Figure 2f).

2.1.2. Modification to the Sublinear Relationship

As earlier mentioned, some sensors exhibit sublinear response characteristics, which degrade the contrast and blur objects. As shown in Figure 2g, the output magnitude increases slowly and saturates (the slope decreases constantly) as the stimulus intensity increases. This problem is common for pressure sensors and can be solved by transferring the sensor outputs to a ring oscillator. The voltage signals from the pressure sensors are transformed into pulse trains that encode the intensity information as a frequency.^[31,32] Conversely, the direct modification of the sublinear response characteristics of the sensors is more efficient, which can increase the effective dynamic range. Therefore, the responsivity should be decreased and current saturation should be avoided under high-intensity stimuli. A similar effect has been reported in field-effect transistors (FETs), where a high gate bias induces electron injection from the channel to the gate dielectric. For example, in n-type FETs (Figure 2h), the accumulated negative charges in the gate dielectric deteriorate the gate control, leading to a decrease in the channel surface potential, and consequently, the channel current (Figure 2i). Light illumination can also induce high carrier concentration. Therefore, the photoresponsivity can be decreased under high-intensity light illumination through a gate bias. Compared with electron injection from the channel to gate dielectrics, charge transfer between the channel and interfacial traps occurs under a modest gate bias.^[33,34] He et al. reported an organic phototransistor consisting of two complementary bulk heterojunctions (BHJs).^[35] Under light illumination, one BHJ served as the photoactive layer and generated a substantial number of carriers. For another BHJ, the photoexcited carriers were trapped in the gate dielectrics, shielding the gate voltage. Therefore, the channel current decreased with light illumination. With a higher light intensity or longer duration, the degree of gate field shielding was higher, leading to stronger inhibition of the photocurrent, as shown in Figure 2i. This light adaptation characteristic contributed to feature enhancement. Liao et al. reported similar adaptive characteristics in phototransistors based on 2D materials.^[36] Under light illumination, a substantial number of carriers were generated in the MoS₂ channel. By applying a positive gate bias, electrons were trapped by the interfacial defects and the gate voltage was shielded. Therefore, the channel current gradually decreased with light illumination.

2.2. Sensitization

Pain perception is an important function of the sensory system that prevents potential or actual tissue injury. The essential nociceptor component realizes the important functions of threshold, relaxation, allodynia, and hyperalgesia, which can be efficiently realized through diffusive memristors.^[37,38] The diffusive memristor receives the output from the sensor (such as thermal and pressure) and executes the computing functions. When the input stimulus intensity is below a certain threshold (harmless stimuli), the nociceptor exhibits no response. Once the stimuli are above the threshold (harmful stimuli), the nociceptor not only reduces its threshold (allodynia) but also

exhibits a higher response (hyperalgesia), meaning the nociceptor increases its responsivity. After an injury, the nociceptor gradually relaxes to its original state, recovering its threshold and responsivity. This characteristic indicates that nociceptors can heal themselves to function normally instead of adhering to a sensitive injury state. In contrast to electrical implementations that require signal transformation and transfer, sensors with in situ nociceptive characteristics are desirable for fast and energy-efficient pain perception.

The space-charge-limited current bears an intrinsic similarity to nociceptive characteristics.^[39,40] Using trap-rich materials as the functional layer, the sensors can exhibit the essential functions of threshold, relaxation, allodynia, and hyperalgesia. As shown in Figure 3a, the excited carriers first fill the traps without contributing to the current. As the intensity increases and the traps are completely filled, external stimuli can induce a notable current increase. The basic threshold function is well emulated through carrier trapping. Because a higher intensity induces more carriers, the resultant current increases with the stimulus intensity, indicating a higher degree of pain.

Kumar et al. used Sb-doped SnO₂ as a photoactive layer, in which oxygen vacancies could capture photogenerated carriers.^[40] As shown in Figure 3b, the current significantly increased after the light stimuli exceeded the threshold. Meanwhile, the magnitude of the photocurrent increased with increasing light intensity. The threshold characteristics under light illumination were also consistent with the electrical results, confirming the importance of the traps. Above this threshold, a large fraction of traps was already full. Therefore, low-intensity stimuli could induce a notable current. This characteristic emulated the allodynia function, in which the injured and vulnerable nociceptors decreased the threshold to protect themselves from strong stimuli. Meanwhile, stimuli with the same intensity would induce a higher current than before, which was known as hyperalgesia. After an injury, the light intensity at which the photocurrent started to increase decreased. Compared to the non-injured state, the photocurrent from the injured cells was higher at the same light intensity. The trap-rich Sb-doped SnO₂ had response characteristics similar to those of nociceptors. The relaxation function could also be realized through trap dynamics. After the external stimuli were removed, the traps gradually released the electrons/holes, causing the nociceptor to relax back to its original state. Notably, some optical nociceptive studies defined a threshold by hand, which required additional circuits.^[41–43] With trap dynamics, the essential nociceptive functions can be well emulated in sensors, which significantly reduces complexity and time latency.

Nociceptors can also indirectly sense noxious stimuli. The nociceptor senses heat or pressure, whereas another type of stimuli has a destructive effect. This type of pain perception protects the sensory system from potentially harmful stimuli. Zhou et al. doped PMMA with azobenzene-functionalized Au nanoparticles, which served as the functional layer of a memristor.^[44] The memristor exhibited threshold switching behaviors with trap dynamics, which formed the basis of its nociceptive characteristics. Meanwhile, once the intensity of UV illumination exceeded a threshold, the memristor changed the threshold and responsivity, which assisted in distinguishing noxious stimuli.

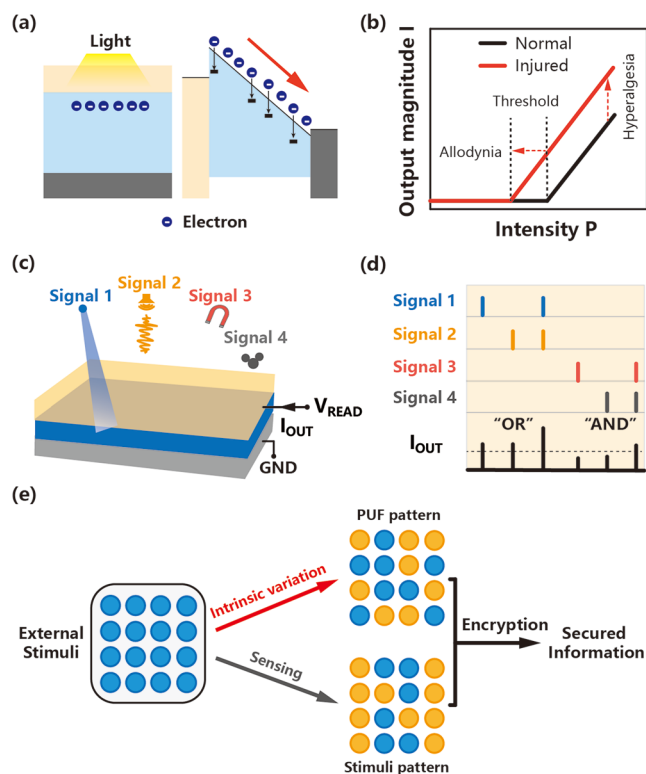


Figure 3. a) Left: Device structure of the artificial nociceptor. The active layer consists of trap-rich semiconductors. b) Under light illumination, the photoexcited carriers first fill in the traps. After the traps are completely filled, the light illumination contributes to notable photocurrent. The photocurrent increases significantly after the light intensity exceeds a threshold. The injured nociceptor exhibits lower threshold and higher responsivity. c) Illustration of in-sensor logic operations. Signals from multiple sources can simultaneously stimulate the sensor. The sources can differ in either the signal types or signal features, such as wavelength, intensity, chemical type. d) Top (Signal 1–4): Different combinations of the signals. Bottom: Output magnitude under different combinations. The dashed line indicates the threshold. e) Schematic of in-sensor PUF design. The sensor array integrates the sensing function with security. The intrinsic variation of characteristic parameters produces a PUF pattern and encrypts the information accepted from the sensor.

2.3. Logic Operation

The logical relationship between different signal sources is important. Logic operations are the functional units of modern digital circuits, which construct important modules including arithmetic/logic units, encoders, decoders, and selectors. Logic operations require multiple inputs, thus multiple electrodes in electrical implementations. In contrast, in-sensor logic functions can mitigate this problem by receiving multiple inputs at the same spatial site, such as light,^[27,45,46] sound,^[47] magnetic fields,^[48,49] and chemicals,^[50,51] as shown in Figure 3c. The input pulses with different features (e.g., wavelength, intensity, and chemical type) represent stimuli from different sources. They stimulate the sensor simultaneously and induce different current magnitudes. The currents are summed inside the sensor, which execute the integration function. The detailed logic operation depends on the response characteristics and threshold, as shown in Figure 3d. Signals 1 and 2 induce output currents higher than

the threshold, which construct the “OR” operation. Signals 3 and 4 induce output currents lower than the threshold; however, the simultaneous input of signals 3 and 4 induces an output current higher than the threshold. Therefore, signals 3 and 4 can construct an “AND” operation. He et al. adopted light pulses of different wavelengths to represent different signal sources.^[52] Short-wavelength light pulses induced a higher photocurrent than long-wavelength light pulses. By setting an appropriate threshold according to the output magnitude, they realized the “OR” and “AND” operations inside the optical sensors.

2.4. Highly Secured Cryptography

An increasing number of sensory devices are being connected to the Internet of Things, generating a large volume of data. Sensory terminals collect audio, images, videos, and other information. Information security has become a significant challenge during the transmission of these data for further computations. To secure information and avoid unauthorized leakage, highly secure algorithms, and hardware should be designed to provide reasonable solutions. A promising approach from the hardware side is a security scheme based on the physical unclonable function (PUF), which is regarded as a promising cryptographic hardware key with irreproducibility and high-level security.^[53,54] It provides a clone-proof device for key authentication and encryption.^[55,56] Conventional PUFs based on silicon complementary metal–oxide–semiconductor (CMOS) technologies are highly integrated and complicated. Owing to the local mismatches of devices and the rapid development of machine learning, it is difficult for such PUFs to achieve a balance between concise design and high robustness. They are weak in handling mathematical attacks because they can be predicted using large-scale machine learning techniques.^[57] Currently, most PUFs are studied as isolated devices in applications only for security, which are still independent of the sensory terminal. The isolated design introduces additional power consumption and physical attack problems. Attackers can disconnect the security scheme from the system and delay the authentication process. Unauthorized individuals may also use reverse engineering techniques to design copycats.

Under these circumstances, security at sensory terminals shows significant potential to overcome conventional drawbacks. Here, we propose a combination device called an in-sensor cryptography device, which requires the realization of sensory functions with a high level of security. The architecture is shown in Figure 3e. Such an aim is challenging based on the design of conventional PUFs.

The development of nanomaterials presents a highly secured alternative to generating robust and tamper-resistant PUFs inside sensors by leveraging the inherent physical disorder and natural randomness of devices.^[58–62] The characteristic parameters, such as the optical response (Raman characteristic peak and responsivity) and electrical response (e.g., mobility μ_{FE} , threshold voltage V_{th} , and subthreshold slope SS) follow a random distribution owing to the defects, moisture, and other uncontrollable factors in the device fabrication process.^[63,64] It meets the basic requirements to generate a PUF. In addition, multiple sensory functions were observed. As earlier mentioned, optical lenses (to capture images), computational units (to

process data), and nociceptors (to accept environmental stimuli) are given as examples. In this case, in-sensor cryptography that authenticates the device at the sensory terminals with security key generation based on its own characteristics can be constructed.

Uniformity, uniqueness, and randomness are the three key criteria to evaluate a PUF system. Both conventional silicon PUFs and novel PUF designs can achieve promising results.^[65] Bit error rate (BER), also known as the reliability that equals $1 - \text{BER}$, is defined to describe the bit error at the output of the PUF device owing to the influence of noise, temperature, and other possible effects. This parameter is introduced to determine the stability of the device. The ideal BER is 0%. Conventional PUFs may encounter a large BER owing to circuit noise and mismatches of the components involved, for example, temperature coefficients or aging effects.^[65] Many novel PUFs based on 2D materials have a lower BER because of their stable characteristics. Theoretically, switchable attributes or characteristic parameters with long-range random distribution can be used to generate keys. Sensors can generate their own PUF keys to authenticate or encrypt data received from the sensing function. For example, we can extract PUF keys based on optoelectronic and electrical parameters.

Oberoi et al.^[64] fabricated an atomically thin and photo-sensitive memtransistor based on monolayer MoS_2 to be used as a PUF device. The Gaussian distribution of its device parameters, such as μ_{FE} , V_{th} , and SS , showed the randomness of such devices. This study generated PUF keys based on I_{DS} when $V_{DS} = 1 \text{ V}$ and $V_{BG} = 0 \text{ V}$. The performance test showed rather good results where the randomness (entropy equals 0.98), uniqueness (inter-hamming distance equals 0.5), and statistical relationship were approximately equal to their ideal value. In addition, the read-out energy was $\approx 1 \text{ nJ}$, as calculated using $\sum I_{DS} V_{DS} \tau_{read}$, where $\tau_{read} = 100 \mu\text{s}$. It was a low power consumption value.^[64] MoS_2 -based PUFs were also good examples.^[63] It provided a concise layout and introduced different types of parameters (optical responses) used for key generation. Meanwhile, numerous studies have revealed the optical active characteristics of MoS_2 , which denoted its ability to capture images. Hence, this device could integrate image sensing and security to realize multiple functions.^[64]

It shows a large potential to merge different types of functions, such as sensing, computing, and storage, with security. Sensory terminals have been realized through different material systems and the variation in inherent characteristics enables these sensors to generate high-performance PUF keys. As a hardware security scheme, PUFs use different entropy sources to provide natural randomness. Various parameters derived from the electrical and optoelectronic characteristics can be used to generate keys (such as resistance, intensity, and dark current). We propose integrating cryptography functions into sensors that can authenticate devices without introducing extra schemes to prevent external physical attacks and reduce latency and energy consumption.

3. In-Sensor Computing at the Array Level

In a conventional sensory system, the sensor array size is much larger than the scale of back-end processing circuits.

The transmission bus and back-end electronic circuits adopt a multiplexing strategy to increase the processing capability, which leads to a tradeoff between the hardware resources and working speed. However, serial signal transmission and processing still increase the time latency. Because the computational sensor can simultaneously execute information processing during signal conversion, highly parallel computations can be realized in the sensor array. This high parallelism arises from the inherent computing ability of the array elements. In addition, conventional sensors work independently and transfer their respective outputs to back-end processing units. To avoid data interference, the sensors do not interact. However, physical coupling between sensors can lead to unique computational functions, even for noncomputational sensors. For example, by applying a read voltage to an interconnected sensor array, all sensor output currents are summed based on Kirchhoff's laws, which represent the spatial average. More complex computing functions can be realized by constructing an interconnected computational sensor array. In this section, we explore two types of computing functions through physical coupling between sensory elements. The response characteristics of the individual sensors, their interaction, and the array structure are discussed in detail.

3.1. Receptive Field

Computational models of biological vision have always been a significant source of inspiration for the design of computer vision algorithms. Small receptive fields in the primary visual system function as fundamental processing units that locally filter a given property of the image according to the spatiotemporal structure of the receptive field. They extract features in the front-end and transfer them to deep hierarchical neural networks for further computing processes. The visual receptive field with a center-surround structure is often identified as the elementary region of the retina. In the two concentric circles of the receptive field, the off-center and on-center retinal ganglion cells have different optical response characteristics. One responds to dark spots surrounded by light backgrounds, whereas the other responds to light spots surrounded by dark backgrounds. This specific spatial alignment of retinal ganglion cells enables feature filtering functions, such as edge and motion detection.^[66] Edge detection is a technique used to identify abrupt changes in image intensity and is often used as a basic building block for more complex algorithms. These changes can be detected using first- or second-order derivatives. Edge detection aims to simplify image processing and accelerate visual perception. It is important in object tracking in self-driving vehicles, drones, or robots.^[67] In digital image processing, the first derivatives of pixel changes for 2D grayscale images are defined according to Equation (1):

$$\Delta f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right] \quad (1)$$

where f is the pixel intensity of point (x, y) . A discrete gradient can be expressed by calculating the difference in current pixels along the partial differential direction according to Equation (2):

$$f(x+1, y) - f(x, y) \text{ and } f(x, y+1) - f(x, y) \quad (2)$$

In this case, we simply add and subtract by convolving the image with filter operators to obtain the gradient for discrete pixels.^[68] Basic edge detection filters include Sobel, Prewitt, and Roberts operators (Figure 4a).^[69] The receptive field is characterized by a higher-level operator and the difference-of-Gaussian kernel, allowing the extraction of edge and contrast information.^[70]

To realize edge detection, optoelectronic sensors should be designed to have positive and negative light responses, corresponding to the positive and negative weights in the convolution operator. A positive photoresponse is common in various photoconductors, in which device conductance increases under light illumination. To realize a more challenging negative photoresponse, researchers have used trap dynamics to suppress the device conductance.^[71] As shown in Figure 4b,^[72,73] the light illumination increases the channel conductance without an electric field, producing an ON photoresponse. When a positive bias is applied at the gate electrode, the electrical field forces the photoexcited electrons to migrate toward the gate and are trapped in the gate dielectric. The negative charges in the dielectric suppress channel conductance, producing an OFF-photoresponse. Meanwhile, heterojunction-based devices can exhibit opposite photoresponses under light illumination with different wavelengths.^[45,74–76] The composite materials have different response characteristics to light illumination with long and short wavelengths owing to characteristics, such as the bandgap and thermal effect.

Assembling these ON- and OFF-photoresponse devices into a filter kernel array according to a specific convolutional operator enables hardware implementation of edge detection. The photocurrent resulting from the analog multiplication and accumulation (MAC) operation is a dot product of the incident light power and the photoresponsivity vectors, as shown in Figure 4c. The resulting non-volatile photocurrents can be summed using Kirchhoff's current law to obtain the convolution result in the output. Wang et al. designed an artificial receptive field in which the light was switched column-wise to represent a contrast-reversing edge moving from the left to the right side. When the edge reached the boundaries of the ON- and OFF-photoresponse devices, the receptive field exhibited a positive/negative peak in photocurrent variation.^[73] One notable challenge is to practically convolve the entire picture. A selector can be used to control the light input for each pixel and switch the image pixels for the convolution computation each time.

In contrast to the receptive field array with rotational symmetrically arranged ON- and OFF-photoresponse devices, the receptive array with photoresponsive devices along one direction (Figure 4d), corresponds to a sub-region of the concentric receptive field and is direction-sensitive for moving objects. In addition to static image processing, Wang et al. used the short-term memory effects of volatile resistive random access memory (RRAM) devices and realized the direction recognition of motion.^[77] Such feature extraction of dynamic images benefited from the specific spatiotemporal structure inspired by the direction-selective ganglion cells. The imitation of spatiotemporal correlation within the receptive field array relied on the appropriate relaxation process of the response to electrical stimuli for each device.

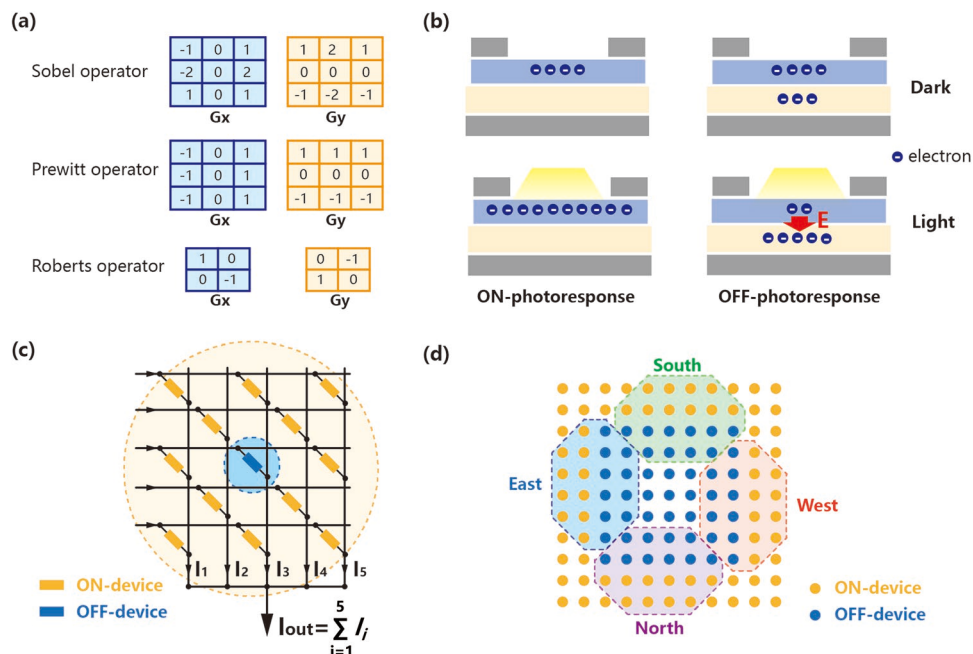


Figure 4. a) Masks for Sobel, Prewitt, and Roberts operators. b) Left: schematic of ON-photoresponse. Right: schematic of the OFF-photoresponse. c) Sketch of the artificial receptive field for edge-detection. The OFF-device is positioned at the center, surrounded by ON-devices. The yellow sphere represents the receptive field of the ON-devices. The blue sphere represents the receptive field of the OFF-device. d) Skeleton of the receptive field for motion detection. The ON/OFF-devices are aligned in one direction.

3.2. Artificial Neural Networks

A sensor array emulating the receptive field can extract important features from external stimuli in situ. Sensors with positive and negative responses can emulate synapses with positive and negative weights, respectively. The computing function of the sensor array is limited by a simple synaptic weight matrix. In-sensor computing also requires high-level sensory processing techniques for more complicated tasks such as recognition, classification, and localization. To realize high-level sensory processing through an in-sensor computing paradigm, reconfigurable sensory devices can be interconnected to directly implement an artificial neural network (ANN) (Figure 5a). Accordingly, biological cognitive processes can be compactly realized in such a single chip without transferring the data outside, thus improving both time and energy efficiency.

Depending on the executed tasks, the sensor array can be specifically designed to implement an ANN. The essential MAC operation using the sensor array is shown in Figure 5b. The inputs to the synaptic array are external stimuli S . The sensors with different weights interact with the external stimuli and output multiplication results, which are summed to obtain the MAC results O_i .

The multiplication process can be executed through the stimuli-electronic conversion process; that is, the multiplication of the stimuli intensity (S) and responsivity (W) results in output currents (O). Figure 5c shows a schematic of the sensor array, in which the dimension of the input stimuli is n and the stimuli contain m classes; that is, the dimensions of the input and output layers are n and m , respectively. Thus, the sensor array consists of n elements (the dashed-line rectangle). Each

element is divided into m sub-elements. All subelements with the same color are physically connected in parallel to sum the resulting currents according to Kirchhoff's current law, $I_m = \sum_{k=1}^n S_k w_{mk}$. Such an operation can be regarded as the inference process of the ANN.

During the training process, the responsivity of each sensor must be continuously modulated controllably. Mennel et al.^[15] designed photodiode arrays based on 2D materials to implement an ANN directly in image sensors. They adopted double local bottom-gate electrodes to electrostatically dope channel materials. Thus, the photoresponse of 2D semiconductors could be increased or decreased by altering the applied gate voltage. Photodiodes simultaneously converted and processed light stimuli without data transfer. The speed of the system was limited only by the optical-electrical conversion process. Therefore, ultrafast image recognition at the nanosecond level could be realized with the rational design of an in-sensor computing architecture.

However, this study was limited to the demonstration of a small-scale network with 27 devices; much is yet to be done prior to the application of this promising technology in large-scale networks or practical applications. First, 2D semiconductors are difficult to uniformly produce over a large area, and it is incompatible with the fabrication process of mature CMOS technology. Second, photodiodes cooperating with the electrostatic doping method lack non-volatile characteristics, which require a continuously applied voltage and cause high power consumption. This problem can be resolved using optoelectronic devices with long-term memory characteristics, for example, floating gate devices.

In addition to the process in which sensors transduce external stimuli to device conductance, the multiplication

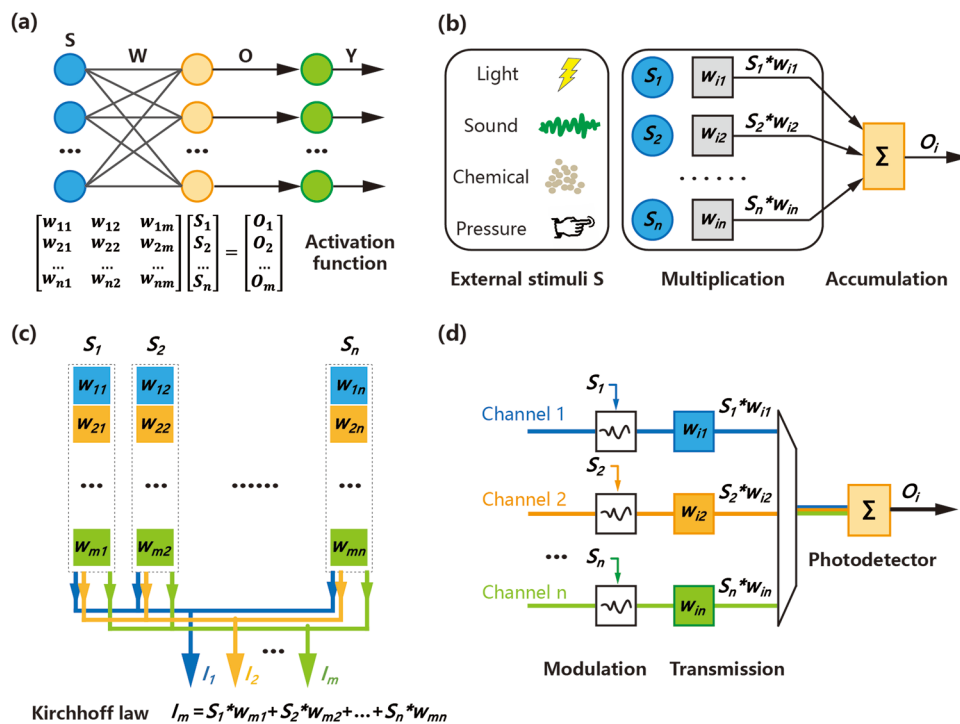


Figure 5. a) Schematic of artificial neural network. S represents the input vector. W represents the synaptic weight matrix. O represents the weighted inputs vector. The bottom panel shows the MAC operation in the synapse array. Y represents the output vector of the activation function. b) Schematic of the MAC operation through the sensor array. The inputs to the array are the external stimuli S . The sensors with different weights interact with the external stimuli and output the multiplication results, which are summed to get the MAC results O_i . c) Schematic of the sensor array in which the sensor responsivity encodes the synaptic weight. The weighted inputs are the output currents of the sensors, which are summed through Kirchhoff's current law. d) Schematic of the sensor array in which the light waveguide encodes the synaptic weight. The weighted inputs are still light signals, which are summed through the accumulation of carriers, electric current, or structural change in the photodetector.

process can also be executed through the light transmission process.^[78–81] As illustrated in Figure 5d, external stimuli modulated the light signal to encode the intensity strength. The light signals in different channels were then transmitted through the waveguide to realize the multiplication operation. The weighted inputs were still light signals, which were summed through the accumulation of carriers, electric currents, or structural changes in the photodetector. Photonics for neuromorphic computing has received much interest because of its ultralow power consumption, high speed, and high parallelism. Frequent and costly MAC operations can be directly implemented in the optical elements. An optical neural network can be integrated with an electronic neural network to complete the entire task by exploiting the advantages of optical and electronic elements. For example, nonlinear activation functions are easier to realize in electronic devices. The overall performance of the sensory system was significantly enhanced by placing optical elements at the front end.

Regarding the scalability of an ANN with an in-sensor computing paradigm, a large neural network with more output nodes significantly increases the complexity of the interconnection between individual sensors. Thus, interconnect routing and layout design require specific consideration. Another important question is the full hardware implementation with online training ability. A specific hardware design is required to realize a training algorithm and the development of hardware-friendly algorithms. However, spiking neural networks provide a

promising alternative solution to enhance efficiency and reduce the footprint through spike time coding.^[82,83] For example, n types of stimuli signals can be encoded as n different spike timings of just one output neuron, which drastically reduces the chip area and hardware implementation complexity. However, hardware demonstrations of in-sensor computing-based spiking neural networks have rarely been reported because of the complexity of implementing spike coding with the sensor device itself. The aforementioned strategy is not limited to the visual system. It can also be extended to other physical stimuli such as auditory, tactile, thermal, or olfactory sensing.

4. Integration Technologies

To satisfy the requirements for the practical application of in-sensor computing, a technology that can integrate large-scale sensor arrays into state-of-the-art silicon technology should be developed. With an increasing need for information processing in different application fields, the development of in-sensor computing is moving toward more specialized information processing units. To realize highly integrated and more energy-efficient processing systems, the hardware implementation of in-sensor computing can adopt a universal and compact layout to fit the existing digital signal processing scheme, which requires a co-design between sensor units, array layout, and peripheral circuits (Figure 6).

For the hardware implementation of the in-sensor computing paradigm, the emerging sensors usually involve materials and structures different from Si-based units, which require the fabrication process of the emerging sensors to be CMOS-compatible. The two-terminal sensor can tune the conductance by intrinsic property changes in the functional layers under environmental stimulations,^[14,84,85] such as MoO_x^[14] and defective SnS.^[85] When two-terminal sensors are assembled into the cross-bar array, the sneak leakage current from neighboring cells leads to operational errors. Thus, a selective cell is required to suppress the leakage current for large-scale integration. A promising solution is to connect a cell selector in series to each sensor, which is inspired by in-memory computing based on RRAM. Compared with a two-terminal sensor, a three-terminal sensor requires an additional gate electrode to tune or retain the conductance of the functional layer.^[36,73,86–88] For such a sensor, a selective cell is unnecessary because weight writing is performed by changing the gate voltage of each sensor. In summary, two-terminal sensors have the advantage of high-density integration, but exhibit disadvantages in terms of reliable performance. In contrast, three-terminal sensors have better controllable performance but inevitable additional power-line problems.

The integration of sensors with selective cells is another major challenge, particularly for large-scale arrays. In practical implementations, the structure of sensors and selective cells should be compatible to achieve overall high performance. In the general design of in-sensor computing, individual sensors must be accessed element-by-element, such as

reading the conductance or programing the characteristics. Conventionally, a transistor can be introduced to construct a one-transistor-one-sensor unit,^[89,90] as shown in the schematic of the array level in Figure 6. Individual or parallel access to sensor cells can be achieved by switching the corresponding transistor ON. In addition to three-terminal transistors, two-terminal selectors with intrinsic nonlinearity have smaller feature sizes that are more suitable for high-density integration.^[91] For the highest integration density, two-terminal sensors and selectors can be serially connected to construct a dense crossbar array, which may result in poor reliability. By replacing the two-terminal selector with a three-terminal transistor, reliability increases at the cost of integration density. The trade-off between performance and integration should be carefully considered.

At the array level, the position and interconnection of the basic sensing units should be carefully designed depending on the desired computing functions. Using feature enhancement as an example, the sensing units are isolated, transferring their respective outputs to the post-processing units. The layout is similar to that of the conventional layout. However, for array-level in-sensor computing, the layouts of the sensing units differ significantly. Unlike the electrical signals directly generated from the peripheral circuits in the RRAM array, the input signals in in-sensor computing are in the form of environmental stimuli (e.g., light, sound, or pressure). As previously illustrated, the layout design should consider the interaction between the environmental stimuli and the sensing units (Figure 6).

For in-sensor computing, the sensor array executes some preliminary computing functions, which reduce some peripheral units, such as current-voltage converters. They are required to bridge the sensor and array in edge computing, which is unnecessary in in-sensor computing. However, array outputs still require complex processing to complete the task. In the case of the receptive field, the sensor array can extract important features such as edges. To complete object recognition, a complex neural network consisting of numerous synapses and neurons is required. The analog outputs are first transferred to an analog-to-digital converter and subsequently to digital processing units and memory (Figure 6). For overall system performance, the back-end processing units and peripheral circuits should be carefully designed. In device-level in-sensor computing, the number of outputs is the same as that of sensory nodes. Therefore, the number of information post-processing units, such as the analog-to-digital converter, also increases when the array size increases. By interconnecting all computing sensors and realizing computation through physical coupling, the number of output nodes decreases slightly. To further reduce the number of postprocessing units, a multiplexing strategy can be adopted. However, this leads to an increase in the information processing time because the processing speed largely depends on the speed of the selective cell and the size of the array. Meanwhile, the shared interconnection between different sensors reduces the resistance variation caused by resistive wires as the fabrication node decreases to the nanometer scale, further improving the computing precision of the overall system. This trade-off requires careful consideration, based on the requirements of the sensory system.

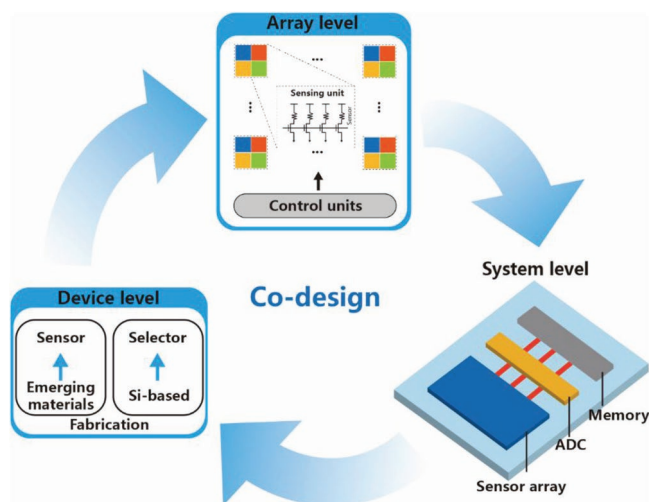


Figure 6. The hardware implementation of in-sensor computing by co-design among device, array, and system. At the device level, the sensors with intrinsic computing ability are primarily based on emerging materials, whereas selector devices are based on silicon. The integration of different materials needs careful investigation. At the array level, the layout depends on the interaction between environmental stimuli and sensing units. The sensing unit consists of sensors and selector devices thus individual access to single sensors is possible. At the system level, since the input signals come from the outside environment, a careful design of signal communication only between the sensors to the post-processing unit is required, for example, the processed sensory information can be directly converted to a digital form by analog-to-digital converter and saved in memory.

5. Conclusions and Perspectives

In-sensor computing is an efficient approach to solving real-time and data-intensive problems. It integrates essential computing functions into the sensor array, which processes information during the transduction of external stimuli into electronic signals. Redundant data are significantly reduced, thereby mitigating the pressure on analog-to-digital signal conversion and data transfer. Therefore, in-sensor computing allows for higher energy efficiency and speed.

The computing functions are derived from two aspects: the device level and array level (summarized in **Table 1**). For in-sensor computing at the device level, the response characteristics of the sensors can be exploited to process information. Depending on the desired computing functions, the response characteristics can vary significantly, which requires careful investigation of materials and devices. We introduce the representative computing functions and explain their basic principles. Feature enhancement aims to amplify the output contrast between stimuli with different intensities. The superlinear relationship between the stimulus intensity and output magnitude produces the best effect, whereas a sublinear relationship degrades the contrast. Nociceptors spontaneously modulate the threshold and responsivity according to the stimulus history. Once the stimulus intensity is above a certain threshold, nociceptors decrease the threshold and responsivity to increase sensitivity. This computing capability also originates from a unique sensing process. Because sensors can collect multiple signals simultaneously, they can process the interaction between them, such as logic or spatiotemporal relationships. Device variations can also be exploited to construct PUFs to endow the sensor with high security. To realize these basic principles, we propose a corresponding physical mechanism and device structure. We mainly discuss the photoelectric effect for clarity; however, the basic principles can be applied to other types of stimuli.

For in-sensor computing at the array level, the inherent computing ability of the sensor renders the array element a computing unit. Because multiple sensors can simultaneously interact with external stimuli, multiple computing operations are executed simultaneously, thereby contributing to high parallelism. In addition, the interconnected structure can provide high computing complexity by exploiting the physical coupling between sensors. The receptive field and ANN can be realized inside the sensor array, significantly reducing the cost of hardware resources, energy consumption, and time latency. In-sensor

computing has proven to be highly promising in this era of big data. This revolutionary change in system architecture provides many benefits. Meanwhile, substantial efforts are also required for materials, devices, integration, fabrication, and algorithms.

The development of in-sensor computing is still in its infancy. Prior experimental studies on in-sensor computing were restricted to specific materials and small arrays. In addition to the general requirements of sensors (e.g., robustness, high dynamic range, and reliability), the device response characteristics should remain stable and predictable. Trap dynamics or electrochemical reactions can induce structural changes and degrade material reliability. To expand the array size for a higher computing ability, the properties of the device should be uniformly distributed. Meeting all these requirements in emerging materials is still challenging; thus, more research on the materials, device structure, and fabrication should be conducted. To apply the in-sensor computing paradigm close to practical applications, utilizing the mature and low-cost silicon technology is efficient, which can significantly reduce the fabrication cost and improve the compatibility with back-end processing units and peripheral circuits, integration density, and reliability. Compared with the conventional design, the in-sensor computing paradigm poses a different requirement for device characteristics. For example, to emulate synapse functions in an ANN, the device response characteristics should be programmable instead of being fixed. The photoresponsivity of the p–n junction-based photodetector can be gradually changed by an external force. By applying a gate bias to electrostatically modulate the doping density, the p–n junction can exhibit different magnitudes of photoresponse, emulating synaptic weight modulation.^[94,95]

Meanwhile, we discuss only a few representative applications. More computing functions and corresponding algorithms are expected, particularly for spatiotemporal patterns. In practice, external stimuli are consistently changing, rendering the precise extraction of spatiotemporal information difficult and costly. For example, in video processing, conventional methods adopt frame-based sampling to obtain multiple images at different times. Subsequently, these images were processed using complex algorithms to extract spatiotemporal information. Because of the large volume of redundant data generated at sensory terminals, many resources are consumed. This can be mitigated by the in-sensor computing paradigm, to a certain extent. For example, dynamic vision sensors only respond to dynamic objects, which significantly reduces data volume. To

Table 1. The summary of in-sensor computing at device level and array level.

	Computing function	Desired characteristics	Ref.
Device level	Feature enhancement	Superlinear relationship between stimulus intensity and output magnitude	[14,20–22,30]
		Modification to sublinear relationship	[35,36]
	Sensitization	Trap dynamics	[39,40,44]
	Logic operation	Sense multiple stimuli simultaneously at the same spatial site	[27,49,52]
	Highly secured cryptography	Entropy sources that provide a random distribution in characteristic parameters	[63,64,92]
Array level	Receptive field	Positive- and negative-response; interconnected sensors	[72,73]
	Artificial neural network	Programmable responsivity, linear response; interconnected sensors	[15,93–95]
		Programmable transmittance; photodetector	[79–81]

process temporal information, memory is essential to store the previous information. Nociceptors process the temporal information on the responsivity changes according to the previous state (injured or normal). The trapped carriers require some time to be fully released, which forms the basis of memory. Therefore, trap dynamics or electrochemical reactions essential to compute functions can also be exploited to process temporal information. The computing capability of the sensory system can be significantly expanded if sensing, computing, and memory can be combined. The benefits of in-memory computing, such as functional polymorphism and information overhead, can also be expected.^[11,12,18,19,72,96–101]

The hardware implementation of sensory systems is a complex and delicate task that involves different materials, device structures, and functional modules. The in-sensor computing paradigm reduces the data transfer between the sensor array and back-end processing units. However, a sensor array utilizing emerging materials may not exhibit high performance in terms of energy consumption and speed. Trap dynamics or redox reactions are much slower than those of conventional devices based on the field effect. Therefore, although some computing functions are already executed by the sensor array, the energy efficiency and speed of the entire system may be restricted by the sensor array. Inspired by in-memory computing, various computing functions must be carefully evaluated to determine the most appropriate ones for in-sensor computing. For example, an RRAM array is suitable for highly parallel and low-precision computing functions. When used as an accelerator for deep learning^[102,103] or combined with a high-precision CMOS processor for scientific computation,^[104] it achieves the best compromise. The same strategy can be applied to in-sensor computing. The precision requirement of the receptive field function is relatively low (positive and negative responses), meaning the non-ideal factors of the sensors (such as reliability, parameter drift, and variation) have a negligible effect on the computing function. Currently, ANNs that require a high integration density and high precision of device responsivity may face challenges in practice.

Acknowledgements

This work was supported by the Research Grant Council of Hong Kong (15205619), the Shenzhen Science and Technology Innovation Commission (SGDX2020110309540000), the Innovation Technology Fund (ITS/047/20), and the Hong Kong Polytechnic University (I-ZETT and CD42).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

bioinspired devices, in-sensor computing, near-sensor computing, optoelectronic devices, vision sensors

Received: April 28, 2022

Revised: June 10, 2022

Published online:

- [1] P. Bellini, P. Nesi, G. Pantaleo, *Appl. Sci.* **2022**, *12*, 1607.
- [2] T. P. Truong, H. T. Le, T. T. Nguyen, *J. Phys.: Conf. Ser.* **2020**, *1432*, 012068.
- [3] W. Ejaz, A. Anpalagan, *Internet of Things for Smart Cities*, 1st ed., Springer, Cham, Switzerland **2019**.
- [4] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, C. Posch, in *2020 IEEE Int. Solid-State Circuits Conf. - (ISSCC)*, IEEE, Piscataway, NJ, USA **2020**, pp. 112–114.
- [5] T.-H. Hsu, Y.-K. Chen, J.-S. Wu, W.-C. Ting, C.-T. Wang, C.-F. Yeh, S.-H. Sie, Y.-R. Chen, R.-S. Liu, C.-C. Lo, K.-T. Tang, M.-F. Chang, C.-C. Hsieh, in *2020 IEEE Int. Solid-State Circuits Conf. - (ISSCC)*, IEEE, Piscataway, NJ, USA **2020**, pp. 110–112.
- [6] P. Lichtsteiner, C. Posch, T. Delbruck, in *2006 IEEE Int. Solid-State Circuits Conf. - Dig. Tech. Pap.*, IEEE, Piscataway, NJ, USA **2006**, pp. 2060–2069.
- [7] B. Wen, K. Boahen, in *2006 IEEE Int. Solid State Circuits Conf. - Dig. Tech. Pap.*, IEEE, Piscataway, NJ, USA **2006**, pp. 2268–2277.
- [8] M. M. Shulaker, G. Hills, R. S. Park, R. T. Howe, K. Saraswat, H. P. Wong, S. Mitra, *Nature* **2017**, *547*, 74.
- [9] Y. Chai, *Nature* **2020**, *579*, 32.
- [10] F. Zhou, Y. Chai, *Nat. Electron.* **2020**, *3*, 664.
- [11] T.-H. Hsu, Y.-C. Chiu, W.-C. Wei, Y.-C. Lo, C.-C. Lo, R.-S. Liu, K.-T. Tang, M.-F. Chang, C.-C. Hsieh, in *2019 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ, USA **2019**, <https://doi.org/10.1109/IEDM19573.2019.8993452>.
- [12] K.-T. Tang, W.-C. Wei, Z.-W. Yeh, T.-H. Hsu, Y.-C. Chiu, C.-X. Xue, Y.-C. Kuo, T.-H. We, M.-S. Ho, C.-C. Lo, R.-S. Liu, C.-C. Hsieh, M.-F. Chang, in *2019 Symp. on VLSI Technology*, IEEE, Piscataway, NJ, USA **2019**, pp. T166–T167.
- [13] T. Wan, S. Ma, F. Liao, L. Fan, Y. Chai, *Sci. China Inf. Sci.* **2021**, *65*, 141401.
- [14] F. Zhou, Z. Zhou, J. Chen, T. H. Choy, J. Wang, N. Zhang, Z. Lin, S. Yu, J. Kang, H. P. Wong, Y. Chai, *Nat. Nanotechnol.* **2019**, *14*, 776.
- [15] L. Mennel, J. Symonowicz, S. Wachter, D. K. Polyushkin, A. J. Molina-Mendoza, T. Mueller, *Nature* **2020**, *579*, 62.
- [16] F. Liao, F. Zhou, Y. Chai, *J. Semicond.* **2021**, *42*, 013105.
- [17] M. Di Ventra, Y. V. Pershin, *Nat. Phys.* **2013**, *9*, 200.
- [18] F. L. Traversa, M. Di Ventra, *IEEE Trans. Neural Networks Learn. Syst.* **2015**, *26*, 2702.
- [19] Y. V. Pershin, M. Di Ventra, in *2014 Int. Workshop on Computational Electronics (IWCE)*, IEEE, Piscataway, NJ, USA **2014**, <https://doi.org/10.1109/IWCE.2014.6865809>.
- [20] R. H. Bube, *J. Appl. Phys.* **1993**, *74*, 5138.
- [21] F. Cardon, R. H. Bube, *J. Appl. Phys.* **1964**, *35*, 3344.
- [22] F. N. Hooge, D. Polder, *J. Phys. Chem. Solids* **1964**, *25*, 977.
- [23] V. Klee, E. Preciado, D. Barroso, A. E. Nguyen, C. Lee, K. J. Erickson, M. Triplett, B. Davis, I. H. Lu, S. Bobek, J. McKinley, J. P. Martinez, J. Mann, A. A. Talin, L. Bartels, F. Leonard, *Nano Lett.* **2015**, *15*, 2612.
- [24] S. Sett, R. S. Bisht, A. Ghatak, A. K. Raychaudhuri, *Appl. Surf. Sci.* **2019**, *497*, 143754.
- [25] C. S. Yang, D. S. Shang, Y. S. Chai, L. Q. Yan, B. G. Shen, Y. Sun, *Phys. Chem. Chem. Phys.* **2017**, *19*, 4190.
- [26] G. Li, D. Xie, H. Zhong, Z. Zhang, X. Fu, Q. Zhou, Q. Li, H. Ni, J. Wang, E. J. Guo, M. He, C. Wang, G. Yang, K. Jin, C. Ge, *Nat. Commun.* **2022**, *13*, 1729.
- [27] T.-Y. Wang, J.-L. Meng, Q.-X. Li, Z.-Y. He, H. Zhu, L. Ji, Q.-Q. Sun, L. Chen, D. W. Zhang, *Nano Energy* **2021**, *89*, 106291.
- [28] F. Zhou, Y. Liu, X. Shen, M. Wang, F. Yuan, Y. Chai, *Adv. Funct. Mater.* **2018**, *28*, 1800080.
- [29] Q. Ma, T. I. Andersen, N. L. Nair, N. M. Gabor, M. Massicotte, C. H. Lui, A. F. Young, W. Fang, K. Watanabe, T. Taniguchi, J. Kong,

- N. Gedik, F. H. L. Koppens, P. Jarillo-Herrero, *Nat. Phys.* **2016**, *12*, 455.
- [30] M. Massicotte, P. Schmidt, F. Vialla, K. Watanabe, T. Taniguchi, K. J. Tielrooij, F. H. Koppens, *Nat. Commun.* **2016**, *7*, 12174.
- [31] Y. Kim, A. Chortos, W. Xu, Y. Liu, J. Y. Oh, D. Son, J. Kang, A. M. Foudeh, C. Zhu, Y. Lee, S. Niu, J. Liu, R. Pfattner, Z. Bao, T.-W. Lee, *Science* **2018**, *360*, 998.
- [32] H. Tan, Q. Tao, I. Pande, S. Majumdar, F. Liu, Y. Zhou, P. O. A. Persson, J. Rosen, S. van Dijken, *Nat. Commun.* **2020**, *11*, 1369.
- [33] Y. Wang, J. Yang, W. Ye, D. She, J. Chen, Z. Lv, V. A. L. Roy, H. Li, K. Zhou, Q. Yang, Y. Zhou, S. T. Han, *Adv. Electron. Mater.* **2019**, *6*, 1900765.
- [34] Z. Lv, M. Chen, F. Qian, V. A. L. Roy, W. Ye, D. She, Y. Wang, Z. X. Xu, Y. Zhou, S. T. Han, *Adv. Funct. Mater.* **2019**, *29*, 1902374.
- [35] Z. He, H. Shen, D. Ye, L. Xiang, W. Zhao, J. Ding, F. Zhang, C.-a. Di, D. Zhu, *Nat. Electron.* **2021**, *4*, 522.
- [36] 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, Y. Chai, *Nat. Electron.* **2022**, *5*, 84.
- [37] J. H. Yoon, Z. Wang, K. M. Kim, H. Wu, V. Ravichandran, Q. Xia, C. S. Hwang, J. J. Yang, *Nat. Commun.* **2018**, *9*, 417.
- [38] G. Ding, R. S. Chen, P. Xie, B. Yang, G. Shang, Y. Liu, L. Gao, W. A. Mo, K. Zhou, S. T. Han, Y. Zhou, *Small* **2022**, *18*, 2200185.
- [39] M. Karbalaee Akbari, J. Hu, F. Verpoort, H. Lu, S. Zhuiykov, *Nano-Micro Lett.* **2020**, *12*, 83.
- [40] M. Kumar, H. S. Kim, J. Kim, *Adv. Mater.* **2019**, *31*, 1900021.
- [41] M. Li, F. S. Yang, H. C. Hsu, W. H. Chen, C. N. Kuo, J. Y. Chen, S. H. Yang, T. H. Yang, C. Y. Lin, Y. Chou, M. P. Lee, Y. M. Chang, Y. C. Yang, K. C. Lee, Y. C. Chou, C. H. Lien, C. L. Lin, Y. P. Chiu, C. S. Lue, S. P. Lin, Y. F. Lin, *Adv. Funct. Mater.* **2020**, *31*, 2007587.
- [42] G. Feng, J. Jiang, Y. Li, D. Xie, B. Tian, Q. Wan, *Adv. Funct. Mater.* **2021**, *31*, 2104327.
- [43] G. Feng, J. Jiang, Y. Zhao, S. Wang, B. Liu, K. Yin, D. Niu, X. Li, Y. Chen, H. Duan, J. Yang, J. He, Y. Gao, Q. Wan, *Adv. Mater.* **2020**, *32*, 1906171.
- [44] L. Zhou, S. R. Zhang, J. Q. Yang, J. Y. Mao, Y. Ren, H. Shan, Z. Xu, Y. Zhou, S. T. Han, *Nanoscale* **2020**, *12*, 1484.
- [45] J. M. Yan, J. S. Ying, M. Y. Yan, Z. C. Wang, S. S. Li, T. W. Chen, G. Y. Gao, F. Liao, H. S. Luo, T. Zhang, Y. Chai, R. K. Zheng, *Adv. Funct. Mater.* **2021**, *31*, 2103982.
- [46] L. Guo, B. Mu, M. Z. Li, B. Yang, R. S. Chen, G. Ding, K. Zhou, Y. Liu, C. C. Kuo, S. T. Han, Y. Zhou, *ACS Appl. Mater. Interfaces* **2021**, *13*, 37645.
- [47] J. Yang, J. Chen, Y. Su, Q. Jing, Z. Li, F. Yi, X. Wen, Z. Wang, Z. L. Wang, *Adv. Mater.* **2015**, *27*, 1316.
- [48] J. Chen, T. Zhang, J. Wang, N. Zhang, W. Ji, S. Zhou, Y. Chai, *Adv. Funct. Mater.* **2021**, *31*, 2104192.
- [49] C. Lv, S. Li, W. Yang, G. Wei, Z. Li, X. Lin, W. Zhao, *IEEE Electron Device Lett.* **2022**, *43*, 482.
- [50] C. Qian, Y. Choi, Y. J. Choi, S. Kim, Y. Y. Choi, D. G. Roe, M. S. Kang, J. Sun, J. H. Cho, *Adv. Mater.* **2020**, *32*, 2002653.
- [51] N. Liu, L. Q. Zhu, P. Feng, C. J. Wan, Y. H. Liu, Y. Shi, Q. Wan, *Sci. Rep.* **2015**, *5*, 18082.
- [52] W. He, Y. Fang, H. Yang, X. Wu, L. He, H. Chen, T. Guo, *J. Mater. Chem. C* **2019**, *7*, 12523.
- [53] B. Shao, T. H. Choy, F. Zhou, J. Chen, C. Wang, Y. J. Park, J.-H. Ahn, Y. Chai, *Nano Res.* **2020**, *14*, 1784.
- [54] C.-H. Chang, Y. Zheng, L. Zhang, *IEEE Circuits Syst. Mag.* **2017**, *17*, 32.
- [55] K. Yang, Q. Dong, D. Blaauw, D. Sylvester, in *2017 IEEE Int. Solid-State Circuits Conf. (ISSCC)*, IEEE, Piscataway, NJ, USA **2017**, pp. 146–147.
- [56] Y. Gao, D. C. Ranasinghe, S. F. Al-Sarawi, O. Kavehei, D. Abbott, *IEEE Access* **2016**, *4*, 61.
- [57] C. Helfmeier, C. Boit, D. Nedospasov, S. Tajik, J.-P. Seifert, in *2014 Design, Automation & Test in Europe Conf. & Exhibition (DATE)*, IEEE, Piscataway, NJ, USA **2014**, <https://doi.org/10.7873/DATE.2014.363>.
- [58] Z. Lin, Y. Zhao, C. Zhou, R. Zhong, X. Wang, Y. H. Tsang, Y. Chai, *Sci. Rep.* **2015**, *5*, 18596.
- [59] J. Guo, C. Liu, Q. Yin, C. Wei, S. Lin, T. B. Hoffman, Y. Zhao, J. H. Edgar, Q. Chen, S. P. Lau, J. Dai, H. Yao, H. S. Wong, Y. Chai, *ACS Nano* **2016**, *10*, 8980.
- [60] J. Yang, Y. Wang, Y. Li, H. Gao, Y. Chai, H. Yao, *J. Mech. Phys. Solids* **2018**, *112*, 157.
- [61] H. Park, A. Afzali, S. J. Han, G. S. Tulevski, A. D. Franklin, J. Tersoff, J. B. Hannon, W. Haensch, *Nat. Nanotechnol.* **2012**, *7*, 87.
- [62] S. Ghatak, A. Nath Pal, A. Ghosh, *ACS Nano* **2011**, *5*, 7707.
- [63] A. Alharbi, D. Armstrong, S. Alharbi, D. Shahrjerdi, *ACS Nano* **2017**, *11*, 12772.
- [64] A. Oberoi, A. Dodda, H. Liu, M. Terrones, S. Das, *ACS Nano* **2021**, *15*, 19815.
- [65] C. Böhm, M. Hofer, *Physical Unclonable Functions in Theory and Practice*, Springer, New York, **2013**.
- [66] R. C. Reid, R. E. Soodak, R. M. Shapley, *J. Neurophysiol.* **1991**, *66*, 505.
- [67] C. Zhan, X. Duan, S. Xu, Z. Song, M. Luo, in *Fourth Int. Conf. on Image and Graphics (ICIG 2007)*, IEEE, Piscataway, NJ, USA **2007**, pp. 519–523.
- [68] P. Dhankhar, N. Sahu, *Int. J. Comput. Sci. Mobile Comput.* **2013**, *2*, 86.
- [69] G. M. H. Amer, A. M. Abushaala, in *2015 2nd World Symp. on Web Applications and Networking (WSWAN)*, IEEE, Piscataway, NJ, USA **2015**, <https://doi.org/10.1109/WSWAN.2015.7210349>.
- [70] C. D. Gilbert, T. N. Wiesel, *Nature* **1992**, *356*, 150.
- [71] H. Nakanishi, K. J. Bishop, B. Kowalczyk, A. Nitzan, E. A. Weiss, K. V. Tretyakov, M. M. Apodaca, R. Klajn, J. F. Stoddart, B. A. Grzybowski, *Nature* **2009**, *460*, 371.
- [72] Z. Zhang, S. Wang, C. Liu, R. Xie, W. Hu, P. Zhou, *Nat. Nanotechnol.* **2022**, *17*, 27.
- [73] C.-Y. Wang, S.-J. Liang, S. Wang, P. Wang, Z. a. 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, F. Miao, *Sci. Adv.* **2020**, *6*, eaba6173.
- [74] H. Li, X. Jiang, W. Ye, H. Zhang, L. Zhou, F. Zhang, D. She, Y. Zhou, S.-T. Han, *Nano Energy* **2019**, *65*, 104000.
- [75] L. Hu, J. Yang, J. Wang, P. Cheng, L. O. Chua, F. Zhuge, *Adv. Funct. Mater.* **2020**, *31*, 2005582.
- [76] C. M. Yang, T. C. Chen, D. Verma, L. J. Li, B. Liu, W. H. Chang, C. S. Lai, *Adv. Funct. Mater.* **2020**, *30*, 2001598.
- [77] W. Wang, E. C. , A. Milozzi, M. Farronato, S. Ricci, C. Sbandati, G. Pedretti, D. Ielmini, *Adv. Intell. Syst.* **2021**, *3*, 2000224.
- [78] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, P. R. Prucnal, *Nat. Photonics* **2021**, *15*, 102.
- [79] B. Shi, N. Calabretta, R. Stabile, *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 7701111.
- [80] H.-T. Peng, M. A. Nahmias, T. F. de Lima, A. N. Tait, B. J. Shastri, *IEEE J. Sel. Top. Quantum Electron.* **2018**, *24*, 6101715.
- [81] J. K. George, A. Mehrabian, R. Amin, J. Meng, T. F. de Lima, A. N. Tait, B. J. Shastri, T. El-Ghazawi, P. R. Prucnal, V. J. Sorger, *Opt. Express* **2019**, *27*, 5181.
- [82] Y. Zhou, Y. Wang, F. Zhuge, J. Guo, S. Ma, J. Wang, Z. Tang, Y. Li, X. Miao, Y. He, Y. Chai, *Adv. Mater.* **2022**, 2107754.
- [83] Y. Zhou, N. Xu, B. Gao, Y. Chen, B. Dong, Y. Li, Y. He, X. S. Miao, in *2019 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ, USA **2019**, <https://doi.org/10.1109/IEDM19573.2019.8993453>.

- [84] J. Lee, S. Pak, Y.-W. Lee, Y. Cho, J. Hong, P. Giraud, H. S. Shin, S. M. Morris, J. I. Sohn, S. Cha, J. M. Kim, *Nat. Commun.* **2017**, *8*, 14734.
- [85] L. Sun, Z. Wang, J. Jiang, Y. Kim, B. Joo, S. Zheng, S. Lee, W. J. Yu, B.-S. Kong, H. Yang, *Sci. Adv.* **2021**, *7*, eabg1455.
- [86] C. Pan, C.-Y. Wang, S.-J. Liang, Y. Wang, T. Cao, P. Wang, C. Wang, S. Wang, B. Cheng, A. Gao, E. Liu, K. Watanabe, T. Taniguchi, F. Miao, *Nat. Electron.* **2020**, *3*, 383.
- [87] S. Wang, C.-Y. Wang, P. Wang, C. Wang, Z.-A. Li, C. Pan, Y. Dai, A. Gao, C. Liu, J. Liu, H. Yang, X. Liu, B. Cheng, K. Chen, Z. Wang, K. Watanabe, T. Taniguchi, S.-J. Liang, F. Miao, *Natl. Sci. Rev.* **2020**, *8*, nwa172.
- [88] T. Y. Qu, Y. Sun, M. L. Chen, Z. B. Liu, Q. B. Zhu, B. W. Wang, T. Y. Zhao, C. Liu, J. Tan, S. Qiu, Q. W. Li, Z. Han, W. Wang, H. M. Cheng, D. M. Sun, *Adv. Mater.* **2020**, *32*, 1907288.
- [89] P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H. S. P. Wong, H. Qian, *Nat. Commun.* **2017**, *8*, 15199.
- [90] C. Li, M. Hu, Y. Li, H. Jiang, N. Ge, E. Montgomery, J. Zhang, W. Song, N. Dávila, C. E. Graves, Z. Li, J. P. Strachan, P. Lin, Z. Wang, M. Barnell, Q. Wu, R. S. Williams, J. J. Yang, Q. Xia, *Nat. Electron.* **2018**, *1*, 52.
- [91] A. Chen, in *2018 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ, USA **2018**, <https://doi.org/10.1109/IEDM.2018.8614505>.
- [92] Z. Hu, J. Comeras, H. Park, J. Tang, A. Afzali, G. S. Tulevski, J. B. Hannon, M. Liehr, S. J. Han, *Nat. Nanotechnol.* **2016**, *11*, 559.
- [93] J. K. Han, Y. W. Chung, J. Sim, J. M. Yu, G. B. Lee, S. H. Kim, Y. K. Choi, *Sci. Rep.* **2022**, *12*, 1818.
- [94] J. Ohta, Y. Nitta, S. Tai, M. Takahashi, K. Kyuma, *J. Lightwave Technol.* **1991**, *9*, 1747.
- [95] K. Kyuma, E. Lange, J. Ohta, A. Hermanns, B. Banish, M. Oita, *Nature* **1994**, *372*, 197.
- [96] B. Chen, F. Cai, J. Zhou, W. Ma, P. Sheridan, W. D. Lu, in *2015 IEEE Int. Electron Devices Meeting (IEDM)*, IEEE, Piscataway, NJ, USA **2015**, <https://doi.org/10.1109/IEDM.2015.7409720>.
- [97] E. Nedaaee Oskoe, M. Sahimi, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2011**, *83*, 031105.
- [98] D. Ielmini, H. S. P. Wong, *Nat. Electron.* **2018**, *1*, 333.
- [99] N. Verma, H. Jia, H. Valavi, Y. Tang, M. Ozatay, L.-Y. Chen, B. Zhang, P. Deaville, *IEEE Solid-State Circuits Mag.* **2019**, *11*, 43.
- [100] F. L. Traversa, M. Di Ventra, *Chaos* **2017**, *27*, 023107.
- [101] Y. V. Pershin, M. Di Ventra, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2011**, *84*, 046703.
- [102] P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, H. Qian, *Nature* **2020**, *577*, 641.
- [103] M. Prezioso, F. Merrih-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, D. B. Strukov, *Nature* **2015**, *521*, 61.
- [104] M. Le Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, E. Eleftheriou, *Nat. Electron.* **2018**, *1*, 246.