SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2026, Vol. 69, Iss. 4, 142404:1–142404:12
https://doi.org/10.1007/s11432-025-4661-y

An optoelectronic artificial spiking neuron array with biomimetic spike-temporal pattern for in-sensor visual prediction

Rui WANG[1,2†], Guolei LIU[2†], Dingwei LI[2†], Xiaotao JING[1], Fanfan LI[2], Zhixian WU[3], Zhongfang ZHANG[2], Huihui REN[2], Saisai WANG[3], Qi HUANG[3], Xiaohua MA[1], Bowen ZHU[2,3,4*], Min QIU[2,3], Hong WANG[1*] & Yue HAO[1]

[1]*State Key Laboratory of Wide Band Gap Semiconductor Devices and Integrated Technology, School of Microelectronics, Xidian University, Xi'an 710071, China*
[2]*Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310024, China*
[3]*Westlake Institute for Optoelectronics, Hangzhou 311421, China*
[4]*Westlake Institute for Advanced Study, Hangzhou 310024, China*

Received 16 May 2025/Revised 8 August 2025/Accepted 1 November 2025/Published online 11 February 2026

Abstract Humans can efficiently predict transient events with a low latency of approximately 80 ms. Simulating this capability to achieve accurate and rapid transient prediction at the edge is essential for advancing bionic machine vision. This task is, however, challenging due to the intricate nonlinear characteristics of dynamic motion, which requires high-throughput data processing and training on digital computers. In this work, we report an optoelectronic artificial neuron (OAN) array to implement in-sensor prediction without training and high-throughput data processing, achieved by emulating spike-temporal patterns of human vision for nonlinear spike encoding and memory. The OAN can nonlinearly encode light intensity into first-spike time (ranging from 7.37 to 0.24 μs) and has a memory ability of 60 s due to oxygen vacancy dynamics. The nonlinear perception and memory capabilities spontaneously generate a spike-recurrent spatiotemporal information equation in situ, allowing the OAN arrays to predict future states using limited data without training. As a result, the array can rapidly and accurately predict pedestrian motion with a high executable frame rate (∼125 fps) and low root mean square error (∼0.014). Combined with spiking neural networks, the array achieves 100% accuracy in pedestrian avoidance. Our research provides forward-looking solutions for advancing bionic machine vision.

Keywords artificial neuron, synapse, spiking neural network, in-sensor computing, prediction

Citation Wang R, Liu G L, Li D W, et al. An optoelectronic artificial spiking neuron array with biomimetic spike-temporal pattern for in-sensor visual prediction. Sci China Inf Sci, 2026, 69(4): 142404, https://doi.org/10.1007/s11432-025-4661-y

1 Introduction

Rapid and accurate prediction of visual dynamic signals at the edge with minimal time cost using limited data is an essential task: it could be of use in a range of applications, including enhancing the safety of autonomous systems and revolutionizing human-computer interaction, particularly in scenarios where decisions must be made within a split second [1–3]. However, current machine vision systems struggle to predict future events in extremely short time frames, due to the complex and dynamic nature of spatiotemporal visual information. Although various methods—such as statistical regression [4], long short-term memory networks (LSTM) [5,6], and reservoir computing (RC) [7,8]—have been deployed on machine vision systems to address this predictability challenge, they all require a large amount of training data and thus fail to reliably predict future evolution from only a short-term time series [9–11]. In addition, current machine vision systems typically comprise hardware with physically separated image sensors and processing units, requiring sensor data to be collected, converted, and transmitted to a digital processor for training [12–14]. The high-throughput iterative problem-solving and high-precision ADCs reduce frame processing rates due to limited data bandwidth, resulting in inefficiencies in energy and time consumption [15]. Therefore, for rapid and accurate motion prediction, it is essential to transfer and fuse visual stream information directly within the sensors to process and predict transient future states.

* Corresponding author (email: zhubowen@westlake.edu.cn, hongwang@xidian.edu.cn)
† These authors contributed equally to this work.

The human visual system is characterized by an exceptionally dense neural network, consisting of over 100 million neurons. This network can process information while signals are being transmitted, enabling rapid and accurate state prediction using limited data [16,17]. This process operates with remarkably low energy consumption, requiring only 20 fJ per operation and functioning with an 80 ms latency [18]. These extraordinary abilities may arise from spike-time patterns in the interactions between neurons and synapses [19]. Neurons exchange information through discrete action potentials, known as "spikes", while synapses serve as the storage units for memory and learning. This facilitates efficient spatiotemporal processing of information before it is transmitted to the brain, culminating in precise predictions within the primary cortex [20, 21].

Developing a vision sensor inspired by biology that seamlessly integrates sensing, spike encoding, spatiotemporal processing, and transient prediction remains a significant challenge. Recent advancements in bio-inspired neurons and synapses have led to promising in-sensor visual hardware capable of dynamic motion processing [22–29]. While these innovations have improved frame processing efficiency and motion detection, the ability to predict transient dynamics directly at the sensor level remains out of reach. Current sensors still rely on extensive data collection and transmission to deep learning frameworks for further training, creating bottlenecks in real-time prediction. Issues such as overfitting on limited data and the need for high-throughput neural network training remain unresolved [30]. These challenges are largely due to the complex interactions between spatiotemporal processes in neurons and synapses, coupled with the absence of simple, biomimetic prediction frameworks that can operate without additional training and be implemented directly within the sensor. Therefore, modern machine visions continue to fall short of achieving the time and energy efficiency of human vision.

In this study, we report an optoelectronic artificial neuron (OAN) array with an In_2O_3 phototransistor (PT) and NbO_x threshold switching (TS) device to replicate the brain's spike-timing patterns for in-sensor visual prediction. The In_2O_3 PT and NbO_x TS replicate the functions of synapses and neurons, respectively. The In_2O_3 PT can sense, store, and learn sensory information, and the NbO_x TS can encode the information into electrical spike signals. Combined with the first-spike time coding rule, the array performs nonlinear and energy-efficient (5.6 pJ per event) spike encoding and storage ability to achieve a neuromorphic spike-temporal pattern, which in situ generates a spike-recurrent spatiotemporal information (RSTI) equation to perceive the future at the sensor level. The OAN arrays have excellent photoresponsivity (60 A/W), rapid spike response (0.24–7.37 μs), and stable spike encoding capabilities over 100 million cycles, laying the foundation for reliable in-sensor visual prediction. As a result, the OAN arrays can accurately predict pedestrian motion with high frame rates (∼125 fps), achieving a root mean square error (RMSE) of ∼0.014—significantly outperforming traditional RC and LSTM models by factors of 68000 and 200, respectively. Combined with spiking neural networks (SNN) for multi-step ahead decision-making, we achieved 100% accuracy in pedestrian obstacle avoidance, compared to only 68% without prediction. This advancement enables sensors to anticipate future events, providing innovative solutions for advancing bionic machine vision.

2 Methods

2.1 Fabrication process of OAN arrays

We first prepare the In_2O_3 synaptic transistor. The preparation process is as follows: the preparation of the In_2O_3 precursor solution (0.1 M) involved dissolving indium nitrate in 2-methoxyethanol. To facilitate the combustion reaction, 0.1 M acetylacetone and ammonium hydroxide solution were added as fuels. After overnight stirring, the precursor solution underwent filtration using a 0.2 μm syringe filter before application. A 2-inch polished silicon wafer with SiO_2 (100 nm) served as the substrate, subjected to sequential cleaning via ultrasonication in Standard Clean 1 (SC-1, NH_4OH:H_2O_2:H_2O, v:v = 1:2:7), deionized water, Standard Clean 2 (SC-2, HCl:H_2O_2:H_2O = 1:1:6), and deionized water, each for 30 min at 70°C. A bottom gate of 10/30 nm Cr/Au was deposited via e-beam evaporation and patterned using the lift-off process. Subsequently, 40 nm Al_2O_3 was deposited through atomic layer deposition (ALD) at 200°C. After 10 min of O_2 plasma surface treatment, the In_2O_3 precursor solution (0.2 M) was spin-coated at 2500 r/min for 40 s (10 nm) and pre-baked at 200°C for 5 min. Following photolithography patterning, the In_2O_3 was etched using HCl:H_2O (1:15, v:v) for 3 s to form the channel patterns. The In_2O_3 channel was then annealed at 300°C for 1 h to form a 10 nm In_2O_3 film. Subsequently, Al_2O_3 was etched using H_3PO_4:H_2O (1:1, v:v) to expose the extended gate electrodes. Finally, 5/50 nm Ni/Au was deposited via e-beam evaporation as the Source/Drain (S/D) electrodes and patterned using the lift-off process, defining a width-to-length ratio (W/L) of 400/10 μm for the phototransistor.

Subsequently, NbO_x threshold switching devices were integrated into the source electrode of the In_2O_3 synaptic transistor following a specific procedure. Initially, UV lithography and lift-off processes were employed, succeeded

by the deposition of a bottom electrode consisting of Ti (5 nm)/Pt (35 nm) onto the source electrode using electron beam evaporation. Subsequently, the non-crystalline NbO_x layer was deposited via magnetron sputtering and patterned using lift-off techniques at room temperature. During the sputtering process, the substrate was cleaned for 60 s using Ar gas at 50 W sputtering power and 8 Pa sputtering pressure. Afterward, a pre-sputtering step was conducted using Ar gas at 50 W sputtering power and 0.5 Pa sputtering pressure. Subsequently, the deposition of a 35 nm NbO_x layer was achieved by sputtering under the same conditions. Following this, the top electrodes, Ti (5 nm)/Pt (35 nm), were deposited via e-beam evaporation and patterned using UV photolithography and lift-off processes to establish the connection between the transistor and NbO_x alignment. The electrode of NbO_x devices measures 1 μm, with an effective working area of 1 μm×1 μm.

2.2 Characterization of the OAN arrays

The cross-section TEM and energy-dispersive X-ray spectroscopy were analyzed by transmission electron microscopy (FEI Tecnai TF-20, UK). The XPS characterization was conducted on X-ray Photoelectron Spectroscopy (ESCALAB Xi+, thermofisher). The OAN arrays was measured using a DC mode by an Agilent B1500 semiconductor parameter analyzer. A B1530A fast measurement unit module was used for generating the voltage pulse and measuring the spiking responses for nonlinear spike spatiotemporal memory encoding. Room-temperature electrical measurements were carried out using a Summit 1100B-M Probe Station.

2.3 Spike-RSTI equation for in-sensor prediction

When confronted with a dynamic visual scene, the spiking neurons array detects a motion denoted as $X^m = (x_1^m, x_2^m, \cdots, x_D^m)$, where D represents the data dimension of the observed motion at time m. Subsequently, it undergoes nonlinear spiking encoding by OAN arrays to yield $T^m = T^m + \gamma T^{m-\tau} = (t_1^m, t_2^m, \cdots, t_D^m)$, expressed as $T = F(X)$, thus establishing the RSTI equation based on the array. Refer to (1) for a detailed representation of this equation,

$$\begin{cases} AF(X^m) = A(T^m + \gamma T^{m-\tau}) = Y^m, \\ F(X^m) = T^m + \gamma T^{m-\tau} = BY^m, \end{cases} \tag{1}$$

where $AB = I$, A is an $L \times D$ matrix, B is a $D \times L$ matrix, I represents an $L \times L$ identity matrix, and γ is the memory factor with $\gamma \geqslant 0$. Expanding the above equation, we can further obtain

$$\begin{cases} A_{L\times D}(t_1, t_2, t_3, \cdots, t_m)_{D\times m} = Y_{L\times m}, \\ B_{D\times L}Y_{L\times m} = (t_1, t_2, t_3, \cdots, t_m)_{D\times m}, \\ A_{L\times D}B_{D\times L} = I_{L\times L}, \end{cases} \tag{2}$$

where $I_{L\times L}$ is the identity matrix. In (3), all of $F(X_m)$ or T_m are available. $A_{L\times D}$ and $B_{D\times L}$ are initially given as null matrices, and $y^{m+1}, y^{m+2}, \cdots, y^{m+L-1}$ are initialized to be 0. By solving the RSTI equations for given T^m, we can obtain the future values $y^{m+1}, y^{m+2}, \cdots, y^{m+L-1}$ of the target variable, as well as the unknown weight matrices A and B. Numerous techniques exist for solving the conjugated equation (4), including the Levenberg-Marquardt method [31]. Within the RSTI equation of the array, information flow undergoes the following transformations: $X^t \to T^t \to Y^t$.

2.4 RMSE calculation

The RMSE is calculated using the following equation (3):

$$\sqrt{1/N \sum_{i=1}^{n} (y_i - f(x_i))^2}, \tag{3}$$

where y_i is the prediction results of the motion and $f(x_i)$ is the desired output.

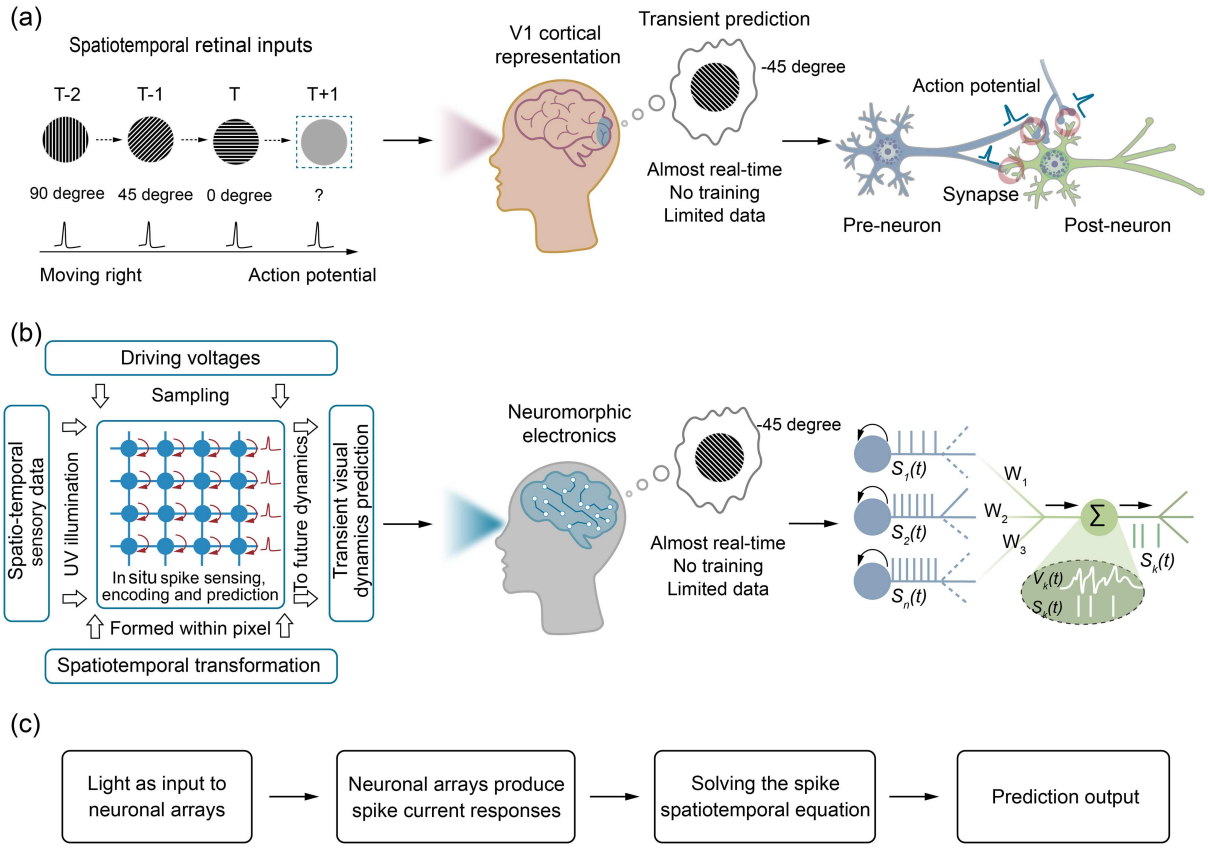

Figure 1 (Color online) Biomimetic in-sensor visual prediction. (a) Illustration of the visual dynamic representation predicted in the primary visual cortex. As an object follows a predictable trajectory, the visual cortex utilizes very short historical data to infer the trajectory and topology of its future motion in nearly real time. (b) Design a neuromorphic vision array, simulating the ability of the brain to predict visual dynamics using limited data. The array spontaneously senses, encodes, learns, and predicts sensing information without training and high-throughput data processing. (c) Flow chat of in-sensor visual prediction.

3 Results

3.1 Biomimetic in-sensor visual prediction

Humans possess intricate internal mental networks within our brains to comprehend the fleeting dynamics of sensory input, and consequently, to forecast trajectories and topologies. This cognitive process is independent of direct sensory input, instead relying on the brain's interpretation of the external environment. Figure 1(a) illustrates a specific scenario, delineating the angular and positional attributes of a small ball undergoing forward rotation at time points $T-2$, $T-1$, and T. Remarkably, the brain anticipates the angle and position of the ball at $T+1$, even with extremely limited data. Recent research indicates that this phenomenon originates in spike-time patterns of the brain [32, 33]. Neurons transduce moving objects into neuronal spikes, while synapses facilitate learning and memory processes. This forms a neural circuit with precise and continuous spiking behavior that can efficiently and comprehensively learn spatiotemporal data. Subsequently, electrical spike signals with spatiotemporal characteristics are transmitted to the primary cortex and are used to accurately predict the future. The capacity of the brain presents an optimal framework for the physical realization of robust neuromorphic vision systems.

In Figure 1(b), we design a neuromorphic vision array to replicate the brain's ability to predict transient vision information. The array takes into account the interaction between optoelectronic synapses and neurons, giving it a neural processing circuit similar to that of biological organisms. When external information is mapped onto the array, the array can encode the information into an electrical spike signal $S(t)$ under the action of the driving voltage, and complete in-situ learning and storage under the action of the synaptic weight W_t. This process produces a spiking equation that allows the array to transform limited present information into a spatiotemporal variable containing future information. Figure 1(c) shows the flow chart of the complete prediction process. Specifically, the whole process can be divided into four steps: light as input to the neuronal array, the neuronal array generating spike current responses, solving the spike spatiotemporal equation, and prediction output. Thus, the array can

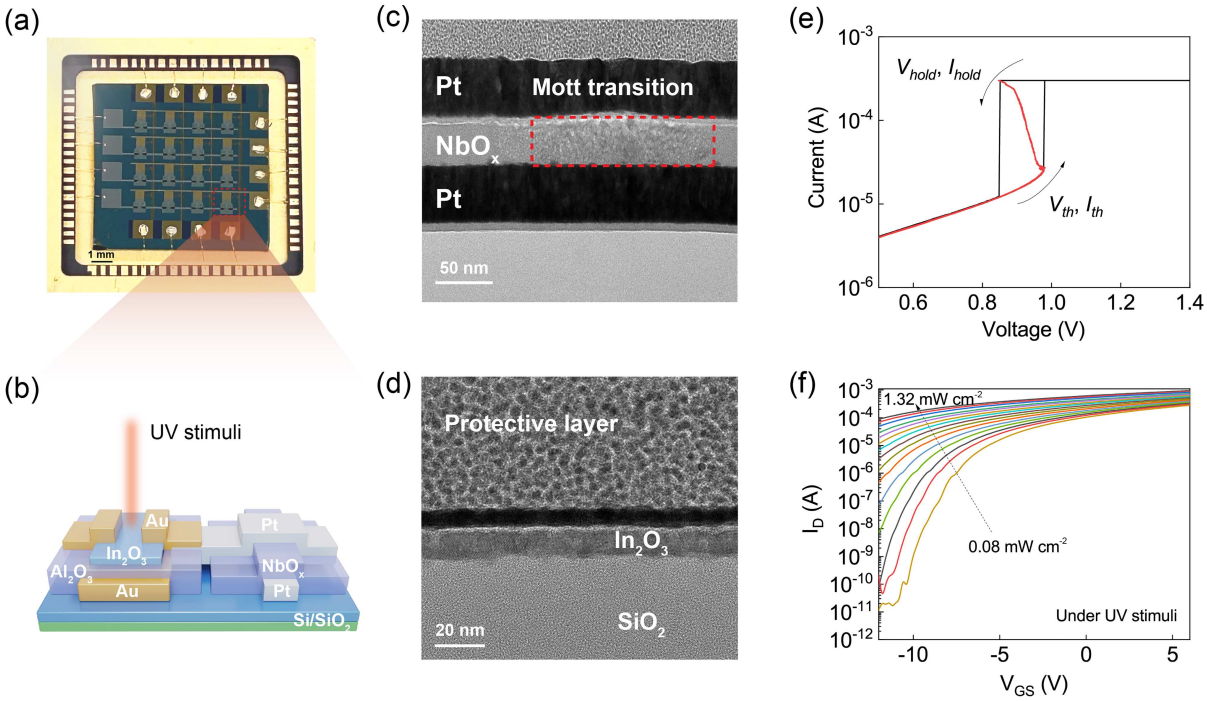

Figure 2 (Color online) Characterization of optoelectronic artificial neuron (OAN) array. (a) Optical images of the OAN arrays. (b) Schematic illustration of integrating In_2O_3 phototransistor (PT) and NbO_x threshold switching (TS) device. (c) Cross-sectional TEM image of NbO_x TS device. Mott transition of NbO_x can simulate the spiking ability of neurons. The red box indicates the Mott transition area. (d) Cross-sectional TEM image of In_2O_3 PT. (e) *I-V* and NDR behavior of the NbO_x TS device. (f) Optoelectronic behaviors of the In_2O_3 PT.

make predictions in near real-time using limited data and without training, just like the primary cortex.

3.2 Optoelectronic artificial spiking neuron array

The neuromorphic vision array was designed by monolithically integrating neurons and optoelectronic synapses. The synapse, composed of In_2O_3 PT, senses, stores, and learns information. The neuron, composed of NbO_x TS devices, undergoes a Mott transition after receiving information from the synaptic transistor, thereby encoding the information into electrical spike signals. In addition, NbO_x TS can naturally replicate biomimetic neuron behavior, while the digital simulation of a single neuron requires hundreds of digital transistors [34]. In_2O_3 PT has advantages such as good thermal stability, high on-current, low operating voltage, and large-area manufacturing, which make a positive contribution to neuromorphic hardware [35]. This forms an OAN arrays consisting of an In_2O_3 PT and NbO_x TS pixel unit. A digital photographic image of the OAN arrays is shown in Figure 2(a). Figures S1(a) and (b) showcase the arrays arranged on a 2-inch wafer and an enlarged photographic image, respectively. Figure 2(b) shows the pixel structure of the array. Digital photographic representations of In_2O_3 and NbO_x devices are provided in Figures S1(c) and (d), respectively. The NbO_x device is integrated into the transistor's source in a sandwich structure with Pt/Ti electrodes, featuring an electrode width of 1 μm. The detailed fabrication process is delineated in Figure S1(c) and Section 2. Figure 2(c) shows the cross-sectional transmission electron microscopy (TEM) image of a fabricated NbO_x TS device. We experimentally observed the Mott transition process of the NbO_x device (shown in the red box). This is triggered by the Mott transition of NbO_2, expected to crystallize at approximately 1000 K when the NbO_x active region maintains a specific stoichiometric ratio (x close to 2) [34]. X-ray photoelectron spectroscopy (XPS) characterization reveals that the Nb:O ratio in the NbO_x film prepared in our experiment remains at 1:2.2 (Figure S2). The energy-dispersive X-ray spectroscopy (EDS) result is shown in Figure S3. The TEM image of the fabricated In_2O_3 PT is presented in Figure 2(d). AFM images and height profiles of the channel region of In_2O_3 film are shown in Figure S4. The thickness of the prepared In_2O_3 film is close to 10 nm. Figure S5 shows the XPS characterization results of In_2O_3, which shows that the oxygen vacancy intensity reaches 0.37, and the resistivity matches well with NbO_x.

The expression of spiking neurons imposes significant demands on the reliability of NbO_x devices. As shown in Figure 2(e), the NbO_x device exhibits a negative differential resistance (NDR) region with TS behavior, where the voltage decreases as the current increases. In this region, the device exhibits Mott phase transition behavior, mimicking the spiking discharge of biological neurons. Figure S6(a) shows the outstanding TS behaviors across 500

current-voltage (I-V) characteristics. Upon surpassing the threshold voltage (V_{th}) of 1.02 V, the device transitions to a low resistance state, returning to a high resistance state when the voltage drops below the hold voltage (V_{hold}) of 0.96 V. Remarkably, no variations are discerned in the distribution of V_{th} and V_{hold} values throughout 500 repetition cycles, and the NbO_x device exhibits repeatable operation for 10^8 cycles, underscoring its exceptional stability and durability (Figures S6(b) and (c)). Transient electrical measurements (triggered by a 1.5 V, 1 μs voltage pulse) reveal a fast response time of ∼10 ns (Figure S6(d)). Figure 2(f) shows the photoelectric behaviors of In_2O_3 PT. The drain current (I_D) exhibits a positive correlation with UV power (P_{in}) at a 365 nm wavelength with a high photoresponsivity of 60 A/W (Figures S7(a) and (b)). The photocurrent (I_{ph}) displays a nonlinear relationship with P_{in} (Figure S7(c)). The non-linearity facilitates in-sensor spatiotemporal processing. In_2O_3 PT has the advantages of low operating voltage, large on-current, and large-area fabrication, making it suitable for monolithic integration with NbO_x for neuromorphic electronics. Figure S8 presents the I-V characteristics of the NbO_x TS and the output characteristics of the In_2O_3 PT. The current levels of the two devices match each other at low operation voltages.

Upon exposure to light, the trapped charge density increases, leading to a negative shift in V_{th}, which is attributed to an increased concentration of ionized oxygen vacancies [36] (Figure S9). This characteristic engenders a persistent photoconductive effect and forms the basis for the spike-temporal pattern. Our preparation of the array achieves a low device-to-device variation (Figure S10). In addition, the threshold voltages of the 16 In_2O_3 phototransistors are shown in Figures S11(a) and (b). The V_{th} of NbO_x and In_2O_3 devices follow a near Gaussian distribution with a mean of 1.17 and −3.03 V, and a standard deviation of 0.07 and 0.34 V, respectively. The mobility of the 16 In_2O_3 phototransistors is shown in Figures S11(c) and (d). The mean and standard deviation are 4.53 $cm^2 \cdot V^{-1} \cdot s^{-1}$ and standard deviation of 0.20 $cm^2 \cdot V^{-1} \cdot s^{-1}$, respectively. In addition, Figure S12(a) shows the R of 16 In_2O_3 phototransistors as a function of V_{GS} under light power of 1.32 $mW \cdot cm^{-2}$. Figure S12(b) shows the statistical analysis of R when V_{GS} is −5 and −3 V, respectively; the mean and standard deviation are 38.4 and 7.7 A/W, respectively, for $V_{GS} = -5$ V and 54.3 and 9.6 A/W, respectively, for $V_{GS} = -3$ V.

3.3 Neuromorphic spike-temporal pattern

Figure 3(a) shows the schematic illustration of the biological spike-temporal pattern. The external stimulation signals are perceived by photoreceptors and converted into electrical spikes by ganglion cells following a neural encoding algorithm, and eventually transmitted to the visual cortex in the brain for processing. After the stimulation is turned off, the firing frequency of neurons gradually decreases rather than disappearing abruptly. The regulation of synapses affects the firing of neurons, resulting in precise and persistent spiking behavior. This allows the brain to capture comprehensive spatiotemporal information. Figure 3(b) shows the implementation of a neuromorphic spike-temporal pattern based on the OAN circuit. By continuous spike stimulation (light and V_{GS}) under bias V_{dd}, the voltage on the transistor decreases until the NbO_x device reaches V_{TH} and switches from HRS to LRS. As a result, the neural circuit discharges and triggers spike current I_{spike}. Due to the persistent photoconductivity effect of the transistor (Figure S13), the NbO_x device can also complete the switching after the light is removed, achieving persistent spike behavior. Detailed explanations are shown in Appendix A.

Figure 3(c) presents the current response of the neuron circuit under different light powers. As the UV power increases from 0.08 to 0.16 $mW \cdot cm^{-2}$, the first-spike time exponentially shifts from 5.78 ± 1.17 to 0.24 ± 0.02 μs. Upon removal of light, the first-spike time gradually diminishes exponentially, enabling a spatiotemporal processing capability. When the light intensity is 0.16 $mW \cdot cm^{-2}$, first-spike time shifts exponentially from 0.24 ± 0.02 to 7.37 ± 0.58 μs as delay time changes from 0 to 60 s (Figure 3(d)). The photoresponse results of different devices are shown in Figures S14 and S15. The error analysis results of the relationship between light power and first-spike time, and between delay time and first-spike time are shown in Figures 3(e) and (f), respectively, which exhibit statistical spiking stability. The persistent and precise spike-firing behavior mimics the brain's spike-temporal patterns and gives neuron circuits short-term working memory capabilities, which can be characterized by a memory factor (γ). The γ is defined by

$$\gamma = T_m / T_{m-\tau}, \tag{4}$$

where T_m is the first-spike time at time m, τ is a sampling interval. The error analysis results of the γ measured at light power of 0.14 $mW \cdot cm^{-2}$ and V_{GS} of 4 V are presented in Figure 3(g). The γ of the three devices are, respectively, 0.41 ± 0.07, 0.43 ± 0.13, and 0.44 ± 0.05, showing stable spike plasticity. The photoresponse and memory enable nonlinear processing capabilities (Figure S16). The nonlinear spike perception and memory lay the foundation for in-sensor prediction.

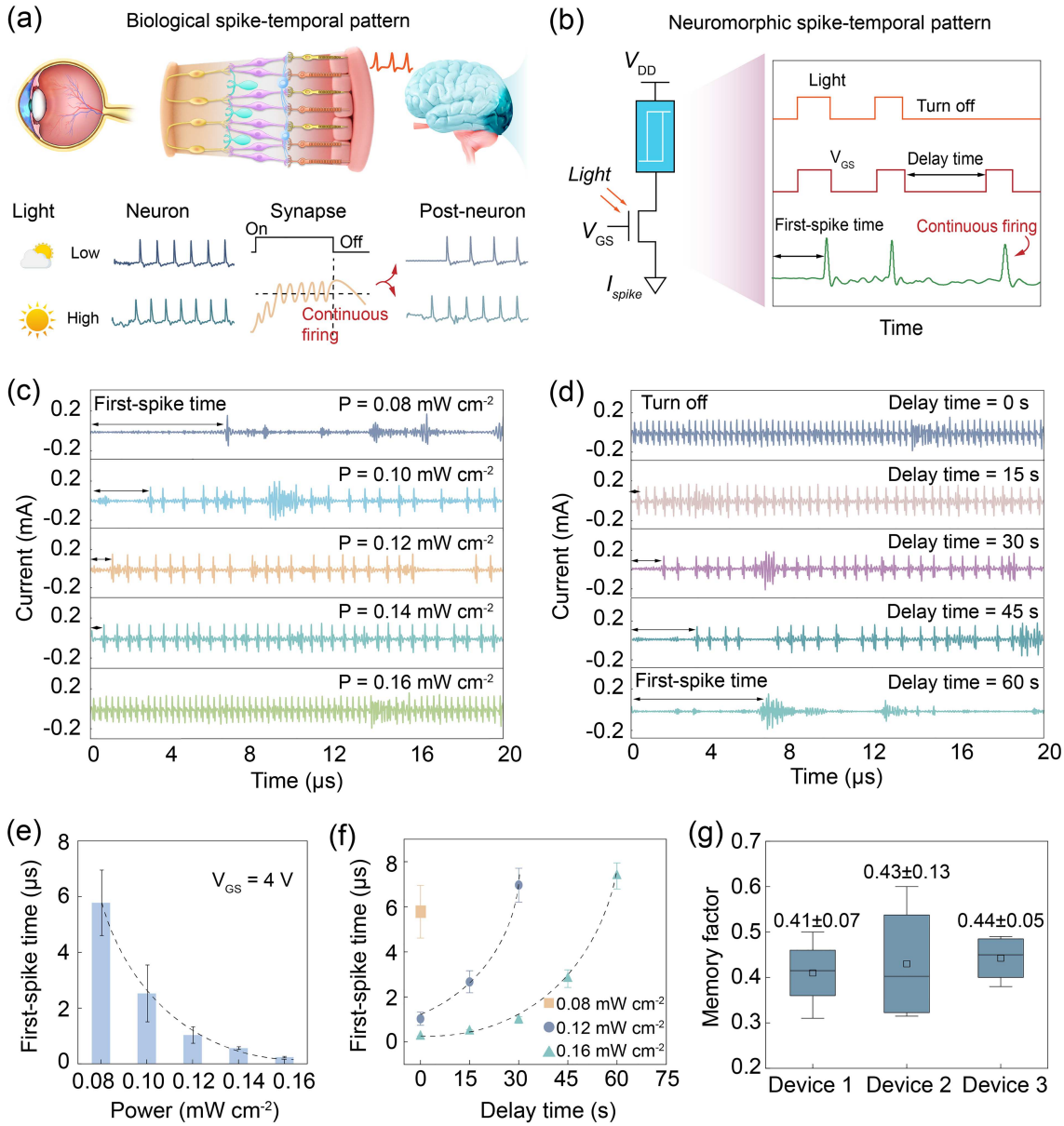

Figure 3 (Color online) Neuromorphic spike-temporal pattern. (a) Schematic illustration of the biological spike-temporal pattern. Neurons have precise and sustained spiking behavior that can capture rich spatiotemporal information. (b) The spike timing pattern is achieved by integrating NbO_x TS devices and In_2O_3 PT, which form an OAN circuit. (c) The output current of the neuron circuit under different light powers at V_{GS} of 4 V. (d) Output current with different delay times after the light stimuli. (e) The first-spike time as a function of light power at V_{GS} of 4 V, showing the precise spiking behavior of the neuron circuit. (f) The first-spike time as a function of delay time under different light powers at V_{GS} of 4 V, showing the sustained spiking memory behavior in the brain. (g) Assessment of sustained spiking memory behavior with the memory factor.

3.4 Experiment implementation of in-sensor prediction

Figure 4(a) illustrates the circuit diagram used to implement in-sensor prediction. The spiking currents ($I^m = I_1^m, I_2^m, \cdots, I_{16}^m$), in the 4 × 4 array at frame m are obtained through row/column addressing, where I_k (k = 1, 2, ⋯, 16) represents a sequence of output currents with a time step of 20 μs. Due to the spike-temporal pattern, the array spontaneously generates an RSTI equation, and the output currents all conform to this equation. In experiments, we found that the future current also satisfies the RSTI equation. Therefore, according to the RSTI equation, we can directly obtain the future current information. Specifically, we applied a V_{GS} pulse (4 V, 20 μs) to the array under UV light (0.14 mW · cm^{-2}) irradiation and then moved the mask to map the light to the array. Here, only 4 frames of data were collected to obtain the RSTI equation for the 5th frame prediction. Each frame of the motion corresponds to a delay time of 15, 30, and 60 s, representing fast, medium, and slow

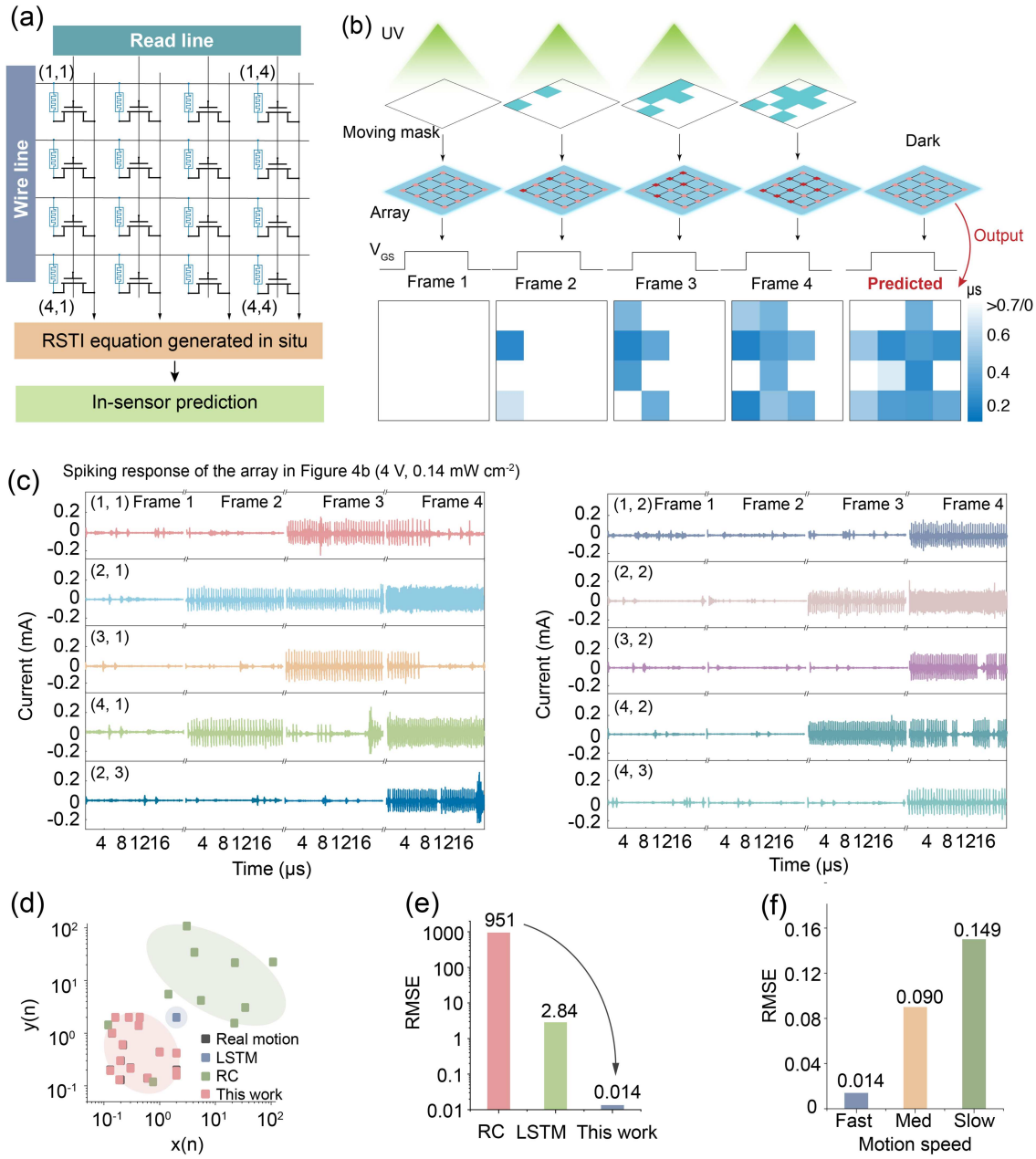

Figure 4 (Color online) Implementation of in-sensor visual prediction. (a) Circuit diagram for the implementation of in-sensor visual prediction. (b) Schematic diagram of the testing process of in-sensor visual prediction. The light with a power of 0.14 mW $\cdot$ cm^{-2} is mapped to the OAN arrays at V_{GS} of 4 V through the 4 $\times$ 4 motion mask (1 to 4 frames), and each mask is mapped to one frame. In the fifth frame, the light and mask are removed at the same time, and the motion information can still be accurately obtained from the array. (c) The array response to moving targets in Figure 4(b). (d) 2-D display of the prediction results with different methods. (e) Our method shows significant performance improvements, with the RMSE reduced by approximately 68000 times and 200 times compared to RC and LSTM, respectively. (f) RMSE values at different speeds.

speeds, respectively. Figure 4(b) shows the operation process and the output of the array for fast motion, where each pixel value denotes the first-spike time of spiking current, extracted from Figure 4(c). The number "0" means no spikes are generated. As a result, the 5th frame, $T^5 = t_1^5, t_2^5, \cdots, t_{16}^5$, was predicted based on the data of the first four frames ($T^m = t_1^m, t_2^m, \cdots, t_{16}^m$, m = 1, 2, 3, 4). The 5th spiking current from the array is depicted in Figure S17. The future motion of the target was faithfully predicted, retaining all its topological attributes. Figure S18 and Section 2 present the theoretical details of the RSTI equation and in-sensor prediction. Figure S19 shows the in-sensor prediction implementation.

RMSE is adopted to quantitatively evaluate the accuracy of the prediction results. Detailed calculations for RMSE are elucidated in Section 2. For comparison, LSTM and RC were also adopted to perform the same task.

Figure 4(d) showcases a 2-D representation of the prediction results obtained from RC, LSTM, and our method. The RMSE for RC was notably high at 951. Despite LSTM yielding a lower RMSE of 2.84, its predictive efficacy was compromised. Conversely, our method demonstrates exceptional performance, boasting a remarkably low RMSE of 0.014, surpassing RC and LSTM by factors of 68000 and 200, respectively (Figure 4(e)). The corresponding RC and LSTM models, as well as the advantages of RSTI, are shown in Appendixes B and C. Figure S20 shows the diagram of the LSTM structure. Figure 4(f) illustrates the prediction results for different movement speeds, showing a gradual increase in RMSE with fast, medium, and slow object speeds: 0.014, 0.09, and 0.149, respectively. The array's predictive ability is influenced by the spatiotemporal plasticity encoding, wherein varying movement speeds entail different γ (Figure S21). Correspondingly, the array's dynamic responses to targets moving at different speeds are presented in Figure S22. As the velocity of object movement decreases, accompanied by a decline in γ, it leads to significant fluctuations in values. The output currents of the first four frames and the 5th frame of medium and slow motion are shown in Figures S23 and S24, respectively.

Further, frame rate, power consumption, and computational complexity are evaluated to demonstrate the great potential for building advanced machine vision. The RSTI equations can be solved in just 8 ms, enabling the array to predict visual motion at approximately 125 fps, exceeding the human visual system's capability of approximately 12 fps (about 80 ms). Since the RDE framework is deployed in the array, this allows the array to establish multiple RSTI equations simultaneously to complete a one-step prediction. Therefore, when the processing unit and the sensor are integrated into a pixel to perform intra-sensor prediction, the prediction time is independent of the array size. Moreover, the power consumption is estimated at approximately 5.6 pJ per event, close to the human neural network's consumption [37] (Figure S25). The total complexity is evaluated at $O(L \cdot n \cdot k \cdot L \cdot L)$, contrasting with LSTM's complexity of $O(L \cdot n_1 \cdot n_2 \cdot N \cdot D + L \cdot s_1^3)$. Given that the observation data requirements are substantially lesser than those of LSTM (i.e., $L \leqslant N, L \leqslant n_2$), the complexity is notably lower than that of LSTM, as detailed in Appendix D.

3.5 Dynamic pedestrian obstacle avoidance

To illustrate the array's potential application in safe and efficient autonomous driving technology, we integrated it with an SNN to implement pedestrian obstacle avoidance. Figure 5(a) illustrates three pedestrian movement scenarios at different velocities: running-fast, jogging-medium, and walking-slow. The vehicle is equipped with the OAN arrays, which predicts the pedestrian's position at time t_{pred} ($t_{pred} = I_{car}/V_{car}$) to facilitate multistep-ahead decision-making. Considering both safety and efficiency, decisions are categorized as follows: if the pedestrian has not entered or has already cleared the road by t_{pred}, the vehicle maintains its current speed. If a pedestrian reaches the other side of the road at t_{pred}, the vehicle decelerates to avert potential hazards from sudden reversals. If pedestrians and vehicles are on the same side of the road at t_{pred}, immediate braking is initiated to ensure a complete cessation of the vehicle within the braking distance. Figures 5(b) and S26 depict the schematic diagram of pedestrian obstacle avoidance.

Pedestrian obstacle avoidance typically encompasses moving obstacle detection, trajectory prediction, and decision-making for obstacle avoidance [38–41]. In our research, the OAN arrays captures dynamic visual data and implements parallel in-sensor prediction. The process of implementing parallel prediction by array is shown in Figure S27 and Appendix E. A CMOS-based active pixel array is demonstrated for pedestrian motion detection, and a comparison between OAN and CMOS active pixel arrays is given in Appendix F and Table F1. To illustrate the advantage of the RSTI equation incorporating bionic spikes, we compare it with the STI equation without OAN arrays encoding. Figure S28 shows the prediction results of pedestrian motion for the future 29 frames. The RSTI equation incorporates bio-inspired spike-time patterns, reducing RMSE by 64.8%. It should be noted that this pulse-time pattern, based on the PPC effect, may introduce cumulative time errors in continuous-time tasks. Figures S29 and S30, respectively, show the motion characteristics, visualization prediction results, and corresponding RMSE values of 9 frames under different memory factors. The results showed that spike timing patterns did enhance the spatiotemporal prediction capabilities of neurons. However, when this effect was too strong, the original motion characteristics were almost lost. When the spike timing pattern had a memory factor of around 0.6, it not only retained the original motion characteristics but also improved the spatiotemporal prediction capabilities of the optoelectronic neuron array. A detailed description is provided in Appendix G. Subsequently, the extracted future features are input into an SNN for multi-step ahead decision-making. Figure 5(c)(top) illustrates both original movement sequences and predicted pedestrian movements for various actions. The 200×200 real scene undergoes extraction into 12×12 features. Only the most salient features per row, sufficient for determining the target's position, are forwarded to the SNN (future 10 frames). Comprising 120 input neurons, 50 hidden neurons, and 3 output neurons, the SNN determines actions—braking, slowing, or maintaining speed. Figure 5(d) depicts the test

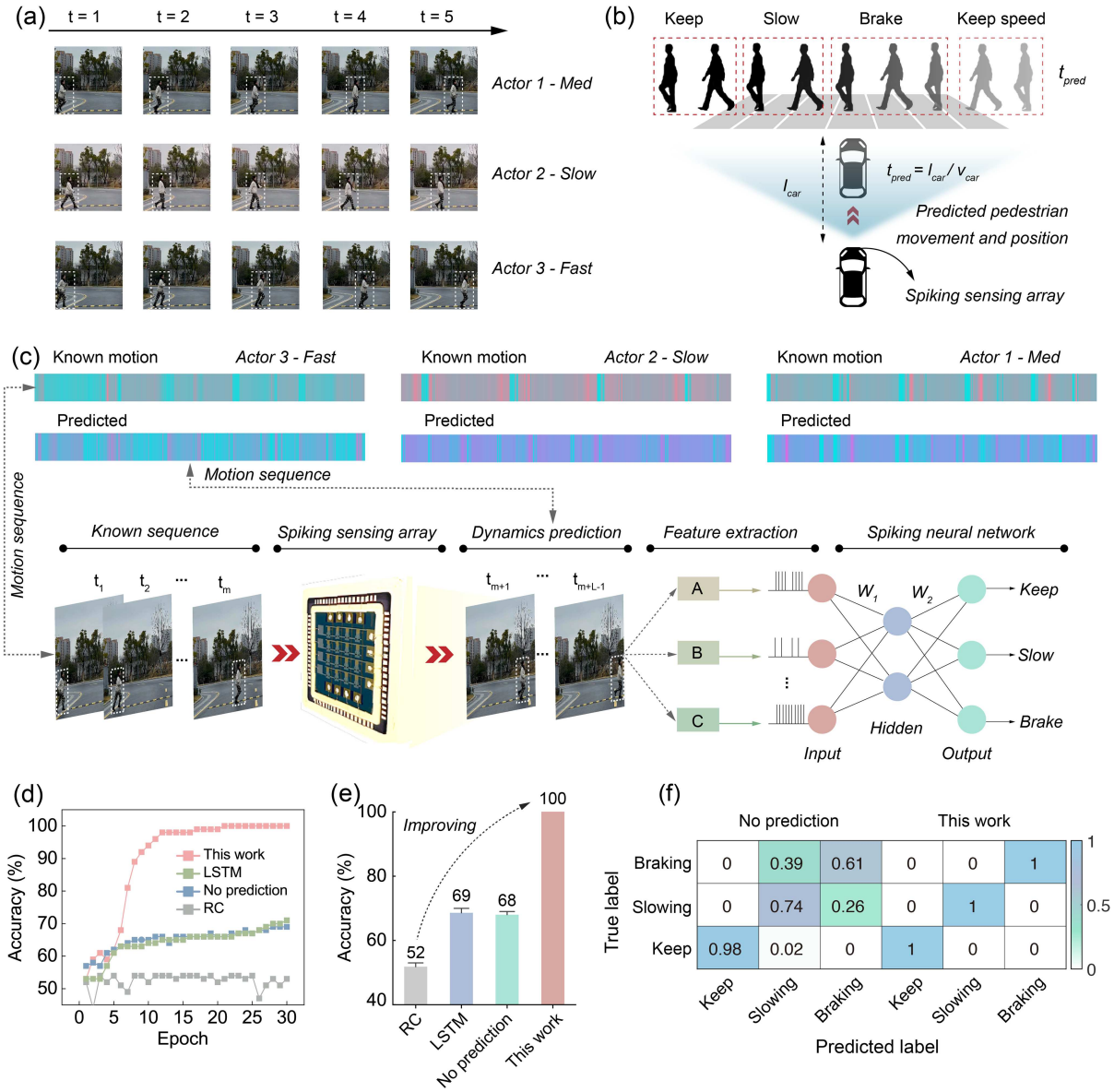

Figure 5 (Color online) Dynamic pedestrian obstacle avoidance. (a) Pedestrians crossing the road at different speeds. (b) A vehicle equipped with the OAN arrays to achieve motion prediction for pedestrian obstacle avoidance. These vehicles forecast the pedestrian's movement and position at $t_{pred} = I_{car}/V_{car}$, allowing for timely decisions such as keeping speed, slowing speed, or braking. (c) Schematic architecture of the pedestrian obstacle avoidance. The array captures and encodes the motion sequence and predicts the future movement and position. Subsequently, the future features are extracted and input into a spiking neural network for multi-step ahead decision-making. (d) Test iteration accuracy for vehicle-pedestrian obstacle avoidance. (e) Comparison of the accuracy of RC, LSTM, and no prediction, our work achieves 100% accuracy in pedestrian obstacle avoidance. (f) Confusion matrix for pedestrian obstacle avoidance.

iteration accuracy concerning vehicle-pedestrian obstacle avoidance. The integration of the array with SNN attains 100% obstacle avoidance accuracy post 20 epochs, as shown in the corresponding accuracy histogram (Figure 5(e)). Figure 5(f) presents a confusion matrix of pedestrian obstacle avoidance decisions. Results indicate that absent prediction, there is approximately a 39% likelihood of pedestrian collision (wherein the vehicle decelerates when braking is necessary), and a nearly 26% probability of opting to brake instead of decelerating, adversely affecting the driving experience. Our method balances safety and efficiency considerations.

In consideration of the intricate real-world environment, susceptible to external interference, we introduced varying degrees of noise (with lengths of 0, 0.5, 1, 1.5, and 2) to real motion data and subsequently fed the predicted results into an SNN for decision-making. Figure S31(a) showcases the test accuracy following 100 epochs of training. Notably, with noise lengths ranging from 0 to 2, the decision accuracy of the active safety system registers at 100%, 99.9%, 98.8%, 97.3%, and 96%, respectively. Corresponding accuracy histograms are depicted in Figure S31(b). Appendix H and Figure S32 provide a detailed SNN implementation. In addition, Figure S33 shows that

the motion intentions of both pedestrians and vehicles can be predicted with high precision by the OAN arrays, verifying its effectiveness in spatiotemporal tracking. Simultaneously predicting car trajectories and pedestrian motion trajectories can help achieve a more comprehensive spatiotemporal processing and prediction system [42]. These findings highlight the great potential of the RSTI equation generated by the array for in-sensor prediction.

4 Conclusion

We have experimentally demonstrated the use of an OAN arrays to perform in-sensor visual prediction without training and high-throughput data processing. The OAN arrays integrates In_2O_3 phototransistors and NbO_x threshold switching devices, which resemble biological synapses and neurons, respectively. With the integration of the first-spike time encoding rule, the OAN arrays can achieve nonlinear, rapid (range from 0.24 to 7.37 μs), and energy-efficient (5.6 pJ per event) spiking encoding and storage, which replicates the spike-temporal patterns in human vision and generates an RSTI equation in situ. In our experiment, we collected the response current generated by the array of moving objects, and all of them satisfied this equation. It is worth noting that the future current response also satisfies this equation, so we can still quickly extract the future current response without external stimulation to implement in-sensor prediction. In addition, the good photoresponsivity (∼60 A/W), and high endurance ($>10^8$) spike encoding and memory behaviors of the OAN arrays have been demonstrated, which enables the array to stably and continuously generate RSTI equations.

We have performed a comprehensive theoretical analysis of the experiments and confirmed the rationality and effectiveness of the in-sensor prediction. Unlike traditional machine vision, which relies on extensive frame collection and high-throughput training, the OAN arrays can achieve a high frame rate (∼125 fps) pedestrian motion prediction using only 4 historical frames, and the prediction accuracy is 68000 and 200 times higher than RC and LSTM methods, respectively. It is worth noting that this predicted frame rate is already better than human vision (∼12 fps). In addition, the future pedestrian motion obtained by the OAN arrays was input into the SNN for advanced decision-making, achieving 100% accuracy in obstacle avoidance, while without prediction, the accuracy was only 68%. It is worth pointing out that due to the good compatibility between In_2O_3 PT and NbO_x TS, spin coating organic matter on In_2O_3 film can expand the spectral range of OAN neurons, thereby broadening the application scope of OAN arrays. Table I1 provides a comparative analysis between our method and other machine vision systems with in-sensor processing techniques, illustrating the substantial advancements this research represents. Overall, our study advances sensing paradigms by endowing arrays with predictive capabilities, providing a forward-looking solution for the evolution of bionic machine vision.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 92464105, 62174138, 62188102) and Key Project of Westlake Institute for Optoelectronics (Grant No. 2023GD004). We thank Westlake Center for Micro/Nano Fabrication, the Instrumentation and Service Center for Physical Sciences (ISCPS), and the Instrumentation and Service Center for Molecular Sciences (ISCMS) at Westlake University for the facility support and technical assistance.

Supporting information Appendixes A–I and Figures S1–S33. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Yang Y, Pan C, Li Y, et al. In-sensor dynamic computing for intelligent machine vision. Nat Electron, 2024, 7: 225–233
2 Chirigati F. Accurate short-term precipitation prediction. Nat Comput Sci, 2021, 1: 709
3 Benigno G B, Budzinski R C, Davis Z W, et al. Waves traveling over a map of visual space can ignite short-term predictions of sensory input. Nat Commun, 2023, 14: 3409
4 Thombs L A, Schucany W R. Bootstrap prediction intervals for autoregression. J Am Stat Assoc, 1990, 85: 486–492
5 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997, 9: 1735–1780
6 Li C, Wang Z, Rao M, et al. Long short-term memory networks in memristor crossbar arrays. Nat Mach Intell, 2019, 1: 49–57
7 Guo X X, Xiang S Y, Zhang Y H, et al. Enhanced memory capacity of a neuromorphic reservoir computing system based on a VCSEL with double optical feedbacks. Sci China Inf Sci, 2020, 63: 160407
8 Kai C, Li P, Yang Y, et al. Human action recognition using a time-delayed photonic reservoir computing. Sci China Inf Sci, 2023, 66: 219401
9 Salmela L, Tsipinakis N, Foi A, et al. Predicting ultrafast nonlinear dynamics in fibre optics with a recurrent neural network. Nat Mach Intell, 2021, 3: 344–354
10 Nishioka D, Tsuchiya T, Namiki W, et al. Edge-of-chaos learning achieved by ion-electron-coupled dynamics in an ion-gating reservoir. Sci Adv, 2022, 8: eade1156
11 Liu K, Zhang T, Dang B, et al. An optoelectronic synapse based on α-In_2Se_3 with controllable temporal dynamics for multimode and multiscale reservoir computing. Nat Electron, 2022, 5: 761–773

12 Chai Y. In-sensor computing for machine vision. Nature, 2020, 579: 32–33

13 Zhong Y, Tang J, Li X, et al. A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing. Nat Electron, 2022, 5: 672–681

14 Bao L, Zhang H S, Wang Z W, et al. Polynomial geometric transformation based on IGZO charge trapping RAM array for machine vision calibration. Sci China Inf Sci, 2025, 68: 160403

15 Mozaffari S, Al-Jarrah O Y, Dianati M, et al. Deep learning-based vehicle behavior prediction for autonomous driving applications: a review. IEEE Trans Intell Transp Syst, 2022, 23: 33–47

16 Petro L S, Muckli L. The brain's predictive prowess revealed in primary visual cortex. Proc Natl Acad Sci USA, 2016, 113: 1124–1125

17 Zou J, He S, Zhang P. Binocular rivalry from invisible patterns. Proc Natl Acad Sci USA, 2016, 113: 8408–8413

18 van Heusden E, Harris A M, Garrido M I, et al. Predictive coding of visual motion in both monocular and binocular human visual processing. J Vis, 2019, 19: 3

19 Fuster J M, Alexander G E. Neuron activity related to short-term memory. Science, 1971, 173: 652–654

20 Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. Nature, 2019, 575: 607–617

21 Chong E, Familiar A M, Shim W M. Reconstructing representations of dynamic visual objects in early visual cortex. Proc Natl Acad Sci USA, 2016, 113: 1453–1458

22 Song H, Lee M G, Kim G, et al. Fully memristive elementary motion detectors for a maneuver prediction. Adv Mater, 2024, 36: e2309708

23 Li F, Li D, Wang C, et al. An artificial visual neuron with multiplexed rate and time-to-first-spike coding. Nat Commun, 2024, 15: 3689

24 Chen J, Zhou Z, Kim B J, et al. Optoelectronic graded neurons for bioinspired in-sensor motion perception. Nat Nanotechnol, 2023, 18: 882–888

25 Tan H, van Dijken S. Dynamic machine vision with retinomorphic photomemristor-reservoir computing. Nat Commun, 2023, 14: 2169

26 Farronato M, Mannocci P, Melegari M, et al. Reservoir computing with charge-trap memory based on a MoS_2 channel for neuromorphic engineering. Adv Mater, 2022,35: e2205381

27 Wang R, Li F, Li D, et al. 1-Phototransistor-1-threshold switching optoelectronic neuron for in-sensor compression via spiking neuron network. In: Proceedings of IEEE International Electron Devices Meeting (IEDM), San Francisco, 2023. 1–4

28 Huang Y J, Tan Y L, Kang Y, et al. Bioinspired sensing-memory-computing integrated vision systems: biomimetic mechanisms, design principles, and applications. Sci China Inf Sci, 2024, 67: 151401

29 Wu Y, Deng W J, Chen X Q, et al. CMOS-compatible retinomorphic Si photodetector for motion detection. Sci China Inf Sci, 2023, 66: 162401

30 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521: 436–444

31 Kanzow C, Yamashita N, Fukushima M. WITHDRAWN: Levenberg-Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J Comput Appl Math, 2005, 173: 321–343

32 Haynes J D, Rees G. Predicting the stream of consciousness from activity in human visual cortex. Curr Biol, 2005, 15: 1301–1307

33 Edwards G, Vetter P, McGruer F, et al. Predictive feedback to V1 dynamically updates with sensory input. Sci Rep, 2017, 7: 16538

34 Kumar S, Williams R S, Wang Z. Third-order nanocircuit elements for neuromorphic engineering. Nature, 2020, 585: 518–523

35 de Wit J H W. The high temperature behavior of In_2O_3. J Solid State Chem, 1975, 13: 192–200

36 Jeon S, Ahn S-E, Song I, et al. Gated three-terminal device architecture to eliminate persistent photoconductivity in oxide semiconductor photosensor arrays. Nat Mater, 2012, 11: 301–305

37 Duan Q, Jing Z, Zou X, et al. Spiking neurons with spatiotemporal dynamics and gain modulation for monolithically integrated memristive neural networks. Nat Commun, 2020, 11: 3399

38 Castillo Aguilar J, Cabrera Carrillo J, Guerra Fernández A, et al. Robust road condition detection system using in-vehicle standard sensors. Sensors, 2015, 15: 32056–32078

39 Gandhi T, Trivedi M M. Pedestrian protection systems: issues, survey, and challenges. IEEE Trans Intell Transp Syst, 2007, 8: 413–430

40 Kim I-H, Bong J-H, Park J, et al. Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques. Sensors, 2017, 17: 1350

41 Choi C, Lee G J, Chang S, et al. Inspiration from visual ecology for advancing multifunctional robotic vision systems: bio-inspired electronic eyes and neuromorphic image sensors. Adv Mater, 2024, 36: e2412252

42 Gao C, Liu D, Xu C, et al. Toward grouped-reservoir computing: organic neuromorphic vertical transistor with distributed reservoir states for efficient recognition and prediction. Nat Commun, 2024, 15: 740