

# Reconfigurable optoelectronic transistors for multimodal recognition

Received: 26 September 2023

Accepted: 5 April 2024

Published online: 16 April 2024

 Check for updates

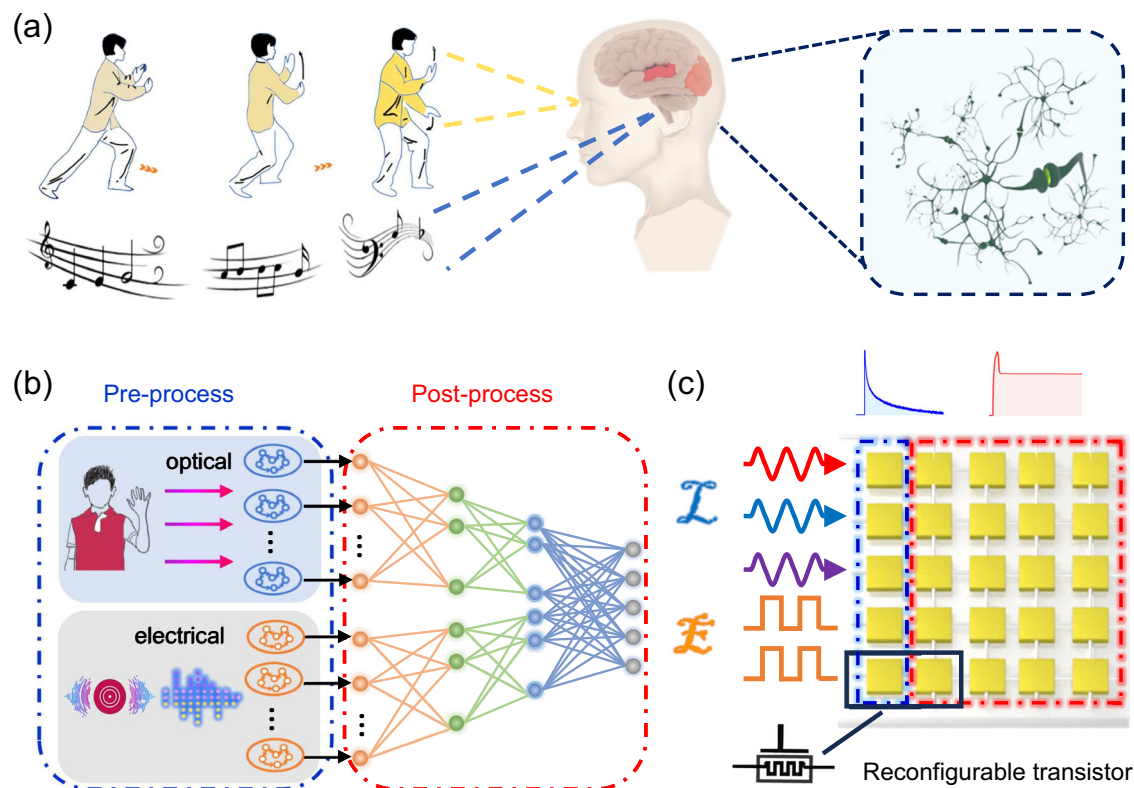
Pengzhan Li<sup>1,2,7</sup>, Mingzhen Zhang<sup>1,3,7</sup>, Qingli Zhou<sup>2,7</sup>, Qinghua Zhang<sup>1,4</sup>, Donggang Xie<sup>1,3</sup>, Ge Li<sup>1,3</sup>, Zhuohui Liu<sup>1,5</sup>, Zheng Wang<sup>1,3</sup>, Erjia Guo<sup>1,3</sup>, Meng He<sup>1</sup>, Can Wang<sup>1,3</sup>, Lin Gu<sup>6</sup>, Guozhen Yang<sup>1</sup>, Kuijuan Jin<sup>1,3</sup> ✉ & Chen Ge<sup>1,3</sup> ✉

Biological nervous system outperforms in both dynamic and static information perception due to their capability to integrate the sensing, memory and processing functions. Reconfigurable neuromorphic transistors, which can be used to emulate different types of biological analogues in a single device, are important for creating compact and efficient neuromorphic computing networks, but their design remains challenging due to the need for opposing physical mechanisms to achieve different functions. Here we report a neuromorphic electrolyte-gated transistor that can be reconfigured to perform physical reservoir and synaptic functions. The device exhibits dynamics with tunable time-scales under optical and electrical stimuli. The nonlinear volatile property is suitable for reservoir computing, which can be used for multimodal pre-processing. The nonvolatility and programmability of the device through ion insertion/extraction achieved via electrolyte gating, which are required to realize synaptic functions, are verified. The device's superior performance in mimicking human perception of dynamic and static multisensory information based on the reconfigurable neuromorphic functions is also demonstrated. The present study provides an exciting paradigm for the realization of multimodal reconfigurable devices and opens an avenue for mimicking biological multisensory fusion.

Humans are bestowed with multi-sensory perceptions and understanding of complex and ever-changing environment<sup>1–3</sup>, and the information obtained through vision and hearing accounts for more than 90% of the total information processed<sup>4,5</sup>. The external dynamic information with dimensional features is perceived and pre-processed by the eyes and ears<sup>6,7</sup>, and then sent to the visual and auditory cortices for post-processing (Fig. 1a). The brain makes decisions and accumulates relevant experience based on the complementary information of the two channels. In this process of continuous experience and

learning, the nervous system plays a vital role, with synaptic plasticity being an important foundation of understanding and adaptation. In contrast, the traditional computer architecture faces the bottleneck of high latency and large energy consumption induced by data shuffling between memory and processing units<sup>8</sup>. Therefore, it is difficult to cope with complex real-world tasks such as machine vision, autonomous driving, and human-machine interaction. Biologically-inspired artificial electronic systems are expected to solve this bottleneck, and have received considerable attention<sup>9–12</sup>. A promising approach is to

<sup>1</sup>Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China. <sup>2</sup>Key Laboratory of Terahertz Optoelectronics, Ministry of Education, Department of Physics, Capital Normal University, Beijing, China. <sup>3</sup>School of Physical Sciences, University of Chinese Academy of Science, Beijing, China. <sup>4</sup>Yangtze River Delta Physics Research Center Co. Ltd., Liyang, China. <sup>5</sup>College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing, China. <sup>6</sup>Beijing National Center for Electron Microscopy and Laboratory of Advanced Materials, Department of Materials Science and Engineering, Tsinghua University, Beijing, China. <sup>7</sup>These authors contributed equally: Pengzhan Li, Mingzhen Zhang, Qingli Zhou. ✉ e-mail: [kjijin@iphy.ac.cn](mailto:kjijin@iphy.ac.cn); [gechen@iphy.ac.cn](mailto:gechen@iphy.ac.cn)



**Fig. 1 | Bioinspired neuromorphic audio-visual fusion system.** **a** Dynamic visual and auditory information obtained via the eyes and ears is transmitted to the relevant areas of the cerebral cortex. Biological neural network integrates two sensory information types to achieve high-level cognition. **b** A neural network architecture diagram for biologically inspired multi-sensory fusion. Left: Reservoir

for preprocessing dynamic information. Right: Artificial neural networks (ANN) for high-level cognition and multimodal information integration. **c** Structural diagram of the hardware implementation of a multimode transistor with reconfigurable properties in response to light and electrical stimuli.

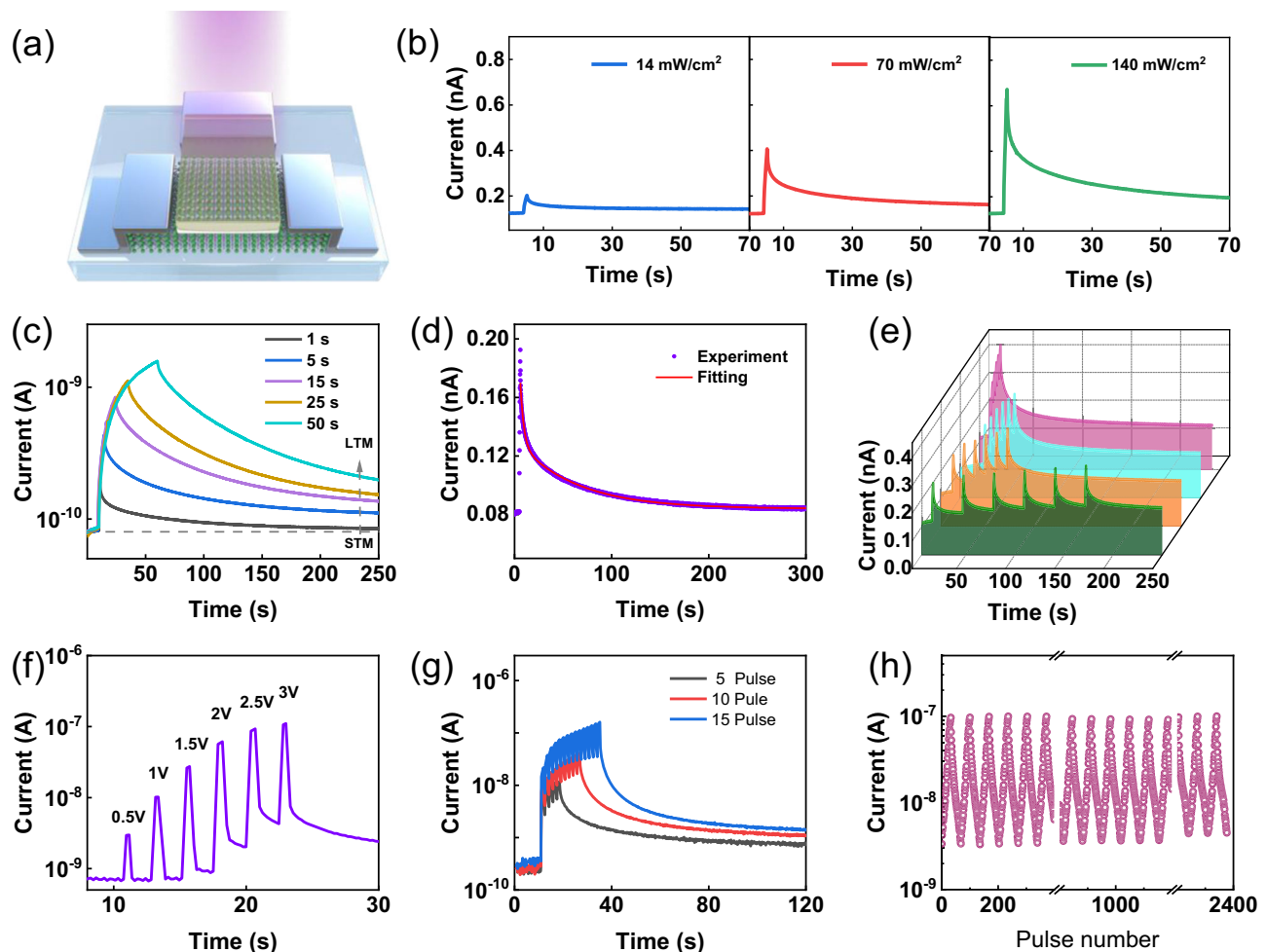
build a compact parallel optoelectronic fusion hardware system and simulate the audio-visual fusion process in the human brain<sup>13–15</sup>.

Multimodal optoelectronic system can be divided into perception/pre-processing and post-processing core, which simulate the functions performed by human receptors and the cerebral cortex, respectively (Fig. 1b). The reservoir is generally used to pre-process input information with sequential characteristics<sup>16,17</sup>. A reservoir relies on nonlinear dynamics to convert low-dimensional signals into a high-dimensional state space<sup>18–20</sup>, thereby enhancing computing efficiency and reducing time and energy consumption<sup>21,22</sup>. The hardware implementation of reservoirs should exhibit volatility, ensuring that the current state of a device is influenced by its recent experience without being affected by distant past events. Subsequently, pre-processed signals are conveyed to an artificial neural network (ANN) for information integration and inference. Artificial synaptic devices can replicate the plasticity of biological synapses, and in the hardware implementation of ANNs, they are responsible for post-processing and storage, necessitating non-volatility<sup>23–25</sup>. Due to the diametrically opposed dynamics required to realize volatile and non-volatile properties<sup>26</sup>, it is difficult to incorporate the two behaviors in a single device. The implementation of the multimodal optoelectronic recognition system shown in Fig. 1b usually requires multiple units with separate reservoir and synaptic functions<sup>27–29</sup>, which greatly increase the complexity of the circuit and integration. In previous studies, researchers generally used ideal software simulation models to replace artificial synaptic devices to perform weight storage and update functions in artificial neural networks<sup>30</sup>, or used sensors to convert different types of information into a single type of electrical signal for follow-up processing<sup>29,31</sup>. And information processing requires additional edge computing devices in most researches. This

essentially limits the potential for future applications of neuromorphic hardware systems. At present, realizing reconfigurable devices with integrated multimode sensing, memory and processing units (Fig. 1c), although desirable, remains a very challenging task<sup>30–32</sup>.

Electrolyte-gated transistors (EGT), with tunable ion dynamic timescales at different gate voltages, show great potential for processing dynamic information in a single device<sup>33,34</sup>. Our previous study demonstrated that the reconfigurable property derives from the electric double layer (EDL) and the ion migration mechanisms<sup>35</sup>. In order to realize multimodal sensing, with this device structure design, BaSnO<sub>3</sub> (BSO) with tunable optical and electrical response was chosen as the channel material. Perovskite-structured BSO has attracted extensive attention due to its wide optical bandgap (~3.1 eV)<sup>36</sup> and high electron mobility at room temperature<sup>37,38</sup>. In addition to its reconfigurable characteristics under electrical regulation, BSO also has temporal and nonlinear memory decay behaviors under ultraviolet (UV)<sup>39</sup> that enables reservoir computing (RC). Therefore, due to its intrinsic properties with tunable multi-timescale optoelectronic response, electrolyte-gated BSO (BSO-EGT) may be a promising building block for multimodal reconfigurable transistors.

In this work, a BSO-EGT that integrates multimodal sensing, memory, and processing functions is introduced. The device can emulate switchable short- and long-term plasticity behaviors under optical and electrical stimulation. Time-scale modulation under UV light exposure originates from the generation of oxygen vacancies, while EDL and ion migration endow reconfigurable properties under voltage stimuli. Thus, both the reservoir and the neural network can be constructed based on the BSO-EGT. This device with multimode sensing and processing capabilities is used to recognize the Fashion-MNIST dataset containing multiple information. Due to its multimodal



**Fig. 2 | EGT with reconfigurable characteristics for multimode sensing.**

**a** Schematic illustration of the neuromorphic transistor that can be stimulated using optical and electrical signals. The BaSnO<sub>3</sub> film serves as a channel between the source (S) and drain (D) electrodes, and IL is used as the gating medium. **b** Short-term potential effect stimulated using optical pulses (1 s) with a 0.3 V bias to monitor the channel current. **c** Transition from short-term memory (STM) to long-term memory (LTM) under different light intensities. **d** Memory decay process after light pulse (70 mW/cm<sup>2</sup> for 1 s) stimulation, fitted using a double-exponential function. The fitting result shows the coexistence of two physical processes.

**e** Excitatory post synaptic current (PSC) induced via a train of optical pulse (70 mW/cm<sup>2</sup>, 1 s with intervals of 1, 5, 20, and 30 s) with a 0.3 V bias to monitor the channel current. **f** The channel current controlled through a series of V<sub>G</sub> pulse with a pulse width of 1 s and different amplitudes, +0.5 V, +1 V, +1.5 V and +2 V, which shows the transition of the PSC from volatile to non-volatile property. **g** Spike number-dependent plasticity under electrical stimuli (+2 V, 1 s) with the same intervals of 1 s. **h** Cyclic controlled LTP using V<sub>G</sub> (equally spaced from +1.5 to +3.5 V, duration of 1 s, spaced 1 s apart) and LTD using V<sub>G</sub> (equally spaced from -1.5 to -3.5 V, duration of 1 s, spaced 1 s apart) for 32 pulses.

nature, the fused information recognition exhibits higher accuracy than that achieved through single-signal processing. We further simulate the function of human audio-visual integration, demonstrating the potential to mimic the superiority of biological multisensory recognition, with an accuracy exceeding 90%.

## Results

### Tunable temporal dynamics under optical stimulation

We epitaxially grew high-quality BSO films on MgO (001) substrates using pulsed laser deposition (PLD). X-ray diffraction of the BSO showed a strong (002) peak accompanied by a (002) peak of MgO due to the epitaxial growth (Supplementary Fig. 1a), and the film thickness was 10 nm determined from the X-ray reflection (Supplementary Fig. 1b). Due to the large lattice mismatch ( $\delta \approx +1.8\%$ ) between the film and substrate, reciprocal space mapping (RSM) around the MgO (204) peak showed that the BSO (lattice parameter,  $a = 0.415$  nm) was relaxed on the MgO ( $a = 0.421$  nm) substrate (Supplementary Fig. 2). High angle annular dark field scanning transmission electron microscopy (HAADF-STEM) and energy dispersive spectroscopy (EDS) results are shown in Supplementary Fig. 3. The atomic-level-resolved

STEM showed a clear interface between the film and substrate, and the BSO film mainly had a perovskite structure<sup>40–42</sup>. The corresponding EDS elemental mapping (Supplementary Fig. 3c–f) indicated the presence and spatial distribution of the BSO film and MgO substrate, and demonstrated a sharp interface. Then, the film was fabricated into an optoelectronic transistor, whose schematic diagram is shown in Fig. 2a. Supplementary Fig. 4 shows the optical microscopy image of the coplanar side-gate BSO-EGT and the left view of the device structure diagram. More details about the fabrication process of the transistor can be found in the “Methods” section.

The channel current  $I_{SD}$  of the device was measured under different UV exposure conditions, and the scanning rate is 9 sampling points/second. The evolution of the channel current over time was monitored via the application of a reading voltage of +0.3 V. With a fixed pulse width of 1 s, the device exhibited volatile and decaying memory behavior when low-intensity light was applied (Fig. 2b). However, as the duration gradually increased, the optical response of the device changed from volatility to non-volatility, similar to the short-term memory (STM) to long-term memory (LTM) transition of biological synapses<sup>43</sup> (Fig. 2c). We also investigated the effect of

different light intensity on the  $I_{SD}$  at the same light duration. As the intensity increased, the conductance exhibited the same transition and decay kinetics (Supplementary Fig. 5). To better describe the decay process of the current, a curve was used to fit the relaxation process after UV illumination, as shown in Fig. 2d. The current descent curve can be fitted well using a double exponential decay equation, as follow:

$$I(t) = C_1 \exp(-t/\tau_1) + C_2 \exp(-t/\tau_2) + C_0 \quad (1)$$

where  $C_0$ ,  $C_1$ , and  $C_2$  are fitting coefficients, and  $\tau_1$  and  $\tau_2$  with fitted values of 1.635 s and 30.418 s are the characteristic time constants, which indicate the coexistence of rapid and slow relaxation in the descent process<sup>44</sup>. The rapid decrease in conductance after UV illumination can be attributed to the electron-hole pair recombination process in BSO<sup>45</sup>, while the slow process is associated with the generation of oxygen vacancies and in-gap states. Oxygen vacancies ( $V_O$ ) generated by UV irradiation are strongly localized on the film surface, which comes from the low mobility of  $V_O$  in BSO films<sup>39,46,47</sup>. It can spontaneously return to the initial state after the stimulus is removed. As the light intensity increases, more oxygen vacancies are generated, leading to the appearance of non-volatility. Based on the above mechanisms, a paired pulse facilitation (PPF) effect can be achieved by applying a pair of UV pulse to the device (Supplementary Fig. 6).

The effect of pulse sequences consisting of eight pulses (pulse width 1 s, light intensity of 70 mW/cm<sup>2</sup>) at different intervals on the device behavior was also studied (Fig. 2e). The results show that shorter pulse intervals result in a higher current rise. The channel current can be stabilized and dynamically varied under the stimulation of a pulse sequence consisting of several pulses at different time intervals. The rise and decay of the channel currents reflect the temporal characteristics of the optical pulse stimulation, which is indicative of the device's high suitability for RC (Supplementary Fig. 7)<sup>34</sup>. Furthermore, the response under blue light (450 nm) illumination was investigated, and a similar volatile behavior as that under low-intensity UV illumination was observed, but with a much smaller magnitude of change (Supplementary Fig. 8). Optical transmittance spectra show strong absorption characteristics at wavelengths shorter than 400 nm (Supplementary Fig. 9a). The BSO films are transparent at visible and near-infrared wavelength due to large band gap of  $E_g \approx 3.1$  eV (Supplementary Fig. 9b). Image recognition and memory are important functions of artificial vision systems. To simulate a UV-sensitive vision system, we fabricated a 3×4 pixelated array using BSO-EGT. Three letters "I", "O" and "P" are used to test the ability of the array to learn and remember images. Under the stimulation of 15 consecutive UV pulses, the image can still be clearly resolved after 350 s, which proves the advantages of our device in simulating the visual system. (Supplementary Fig. 10).

Previous studies have verified that oxygen vacancies will be generated in BSO films under UV irradiation, resulting in an increase of film conductance. Lee et al. used ambient-pressure X-ray photoemission spectroscopy (APXPS) for in situ characterization to monitor the origin of generated defects and proved that the UV illumination under vacuum leads to chemical modification by evolution of oxygen-vacancy-related defects on the surface of BaSnO<sub>3</sub><sup>39</sup>. And the result of cross-sectional scanning transmission electron microscopy (STEM) in vacuum-illuminated BaSnO<sub>3</sub> epitaxial films also confirmed the existence of oxygen-related defects at the surface<sup>39</sup>. When the UV dose [UV dose (mJ/cm<sup>2</sup>) = UV Intensity (mW/cm<sup>2</sup>) × Exposure Time (s)] is low, the oxygen vacancy concentration is low, and its impact on the BSO channel can be recovered quickly. On the other hand, when the UV dose is high, the oxygen vacancy concentration increases, resulting in non-volatile conductance changes in the BSO channel.

## Reconfigurable dynamics of BSO-EGT under electric stimuli

Next, we examined the electrical modulation of BSO through electrolyte gating. Due to the powerful regulation capability of ionic liquids (ILs)<sup>48</sup>, the BSO-EGT was constructed using N,N-diethyl-N-(2-methoxyethyl)-N-methylammoniumbis-(trifluoromethylsulfonyl)-imide (DEME-TFSI) as the electrolyte-gating medium. Supplementary Fig. 11 shows the transistor characteristic curves of BSO-EGT, namely the transfer characteristic curve and the output characteristic curve. Supplementary Fig. 11a illustrates the transfer curves measured along the counterclockwise direction, and the scan rate is 10 mV/s. When the gate bias is +2.5 V, the channel current is 0.75 μA and the leakage current is 0.8 nA. The difference between the two reaches three orders of magnitude, so the effect of leakage current can be neglected. The response of the device to electrical pulse stimulation was then measured at  $V_{SD} = 0.3$  V. At low  $V_G$ , the channel current can quickly decrease to initial state, but the current did not fully recover as the increase of  $V_G$  (Fig. 2f). That is, the device transitioned from volatility to non-volatility with the increase of  $V_G$ , exhibiting typical reconfigurable characteristics. Then, a voltage bias with fixed amplitude (+1 V) and variable pulse width was applied to the gate. The device exhibited volatility, with the current rapidly decaying to initial state when the low  $V_G$  is removed (Supplementary Fig. 12a). During electrical stimulus, the scanning rate is 5 sampling points/second. The volatile response of BSO-EGT at a lower  $V_G$  originated from the rapid movement of anions and cations in ILs under an electric field, which produced a strong accumulation of space charge at the interface between the electrolyte and the channel, called a Helmholtz layer or EDL<sup>49</sup>. Accumulation occurs due to blocking ions in the solid channel. In order to balance the EDL formed at the interface, an accumulation of electrons occurs inside the channel. Therefore, the conductance of the channel will change. When the external bias voltage is removed, the ions at the interface spontaneously migrate back into the ionic liquid, and the EDL disappears, so the channel conductance returns to its original state<sup>35</sup>.

But, under the stimulation of voltage pulses  $V_G = +2$  V with different durations, the conductance of the device can be maintained at a high level without returning to initial state (Supplementary Fig. 12b). Moreover, the correlation of device characteristics with gate voltage pulses of different time intervals, from 0.1 s to 2 s, was tested by applying a pulse sequence (+1 V, 1 s). As the interval decreased, the level of current accumulation became more pronounced (Supplementary Fig. 13). This is due to the fact that only a small part of the ions accumulated at the interface relaxes back into the bulk when the spacing is short, thereby promoting channel doping and conductance. The device conductance can be maintained at different levels as the number of applied pulses increases under  $V_G$  of +2 V (Fig. 2g). To explore the multi-level memory properties of the device,  $I_{SD}$  was adjusted to different conductance levels, and the retention characteristics were tested. There was no significant decay of channel conductance over a period of 300 s (Supplementary Fig. 14), indicating that BSO-EGT has good non-volatility. Furthermore, the pulse-switching characteristics of electrical potentiation (+2 V, 1 s) and depression (-2 V, 1 s) were investigated. The transistor was reversibly switched between the high- and low-conductance states hundreds of times without significant degradation (Supplementary Fig. 15).

The conductance of the BSO channel showed non-volatile changes under high voltage stimulation, and there is a peak in the gate current of the transfer characteristic curve at -1.3 V (Supplementary Fig. 11a), which is related to the hydrolysis reaction<sup>50</sup>. This is because, in addition to the presence of EDL at the interface, Protons ( $H^+$ ) originating from the trace water containing in the ionic liquids would be injected into the film when the positive voltage exceeded the critical voltage. Protons produced by hydrolysis can be driven to the channel interior, resulting in strong interactions with the solid material<sup>49,51</sup>. Therefore, the non-volatility of the device came from the migration of protons generated by hydrolysis in the ionic liquid into the oxide film

under the positive gating, causing a non-volatile increase in the channel conductance.

To verify this mechanism, secondary ion mass spectrometry (SIMS) was performed on the BSO films after applying different electrical stimuli. Supplementary Fig. 16a shows that hydrogen ions appear inside the electrically modulated films, and the hydrogen ion concentration rise significantly as the voltage increased. We added a small volume of D<sub>2</sub>O to the IL, in which D<sup>+</sup> just acts as an isotope marker to show the source of the H<sup>+</sup>. The addition of D<sub>2</sub>O to the IL resulted in the presence of D<sup>+</sup> signal in the electrically-modulated film, which was not observed inside the film without modulation (Supplementary Fig. 16b). Therefore, the non-volatility of the BSO-EGT comes from the injection and diffusion process of hydrogen ions originating from hydrolysis. Additionally, there is a 12-hour waiting time for ionic liquid-gated samples before SIMS experiments were performed. Therefore, this also indirectly reflects that the insertion of hydrogen ions can exist inside the sample for a long time, causing a non-volatile effect. Long-term synaptic plasticity including long-term potentiation (LTP) and long-term depression (LTD) were simulated using our transistor (Fig. 2h). LTP was simulated using 32 consecutive positive electrical pulses (voltages from +1.5 to +3.5 V with a pulse width of 1 s), while LTD appeared when 32 negative voltage pulses were applied to the gate (−1.5 to −3.5 V, the pulse width is 1 s). The above result means that reconfiguration between volatile and non-volatile transistors can be achieved through electrical stimuli by controlling V<sub>G</sub>.

### Multimodal characteristic of optoelectronic BSO-EGT

The temporal dynamics of the BSO-EGT under separate optical and electrical stimulation, provide two means of modulating the device characteristics. Furthermore, the fused optoelectronic response characteristics of the BSO-EGT devices are analyzed. There are three combinations of dual pulse stimulation: two electrical pulses (EE), one light and one electrical pulse (LE), and two light pulses (LL). The evolution of I<sub>SD</sub> under various conditions is shown in Supplementary Fig. 17a, where the width of both optical and electrical pulses is 1 s. The sampling point was the current value after a delay for 1 s from the second pulse. And the scanning rate is 9 sampling points/second. It can be seen that under the three different stimulation modes, the current of the device reaches three distinct states. This is because the application of a light pulse followed by an electrical pulse induces a further increase in conductance, and the two kinds of stimuli have different relaxation dynamics. Therefore, the response to multimodal inputs exhibits different decay characteristics compared to those obtained under individual light or electrical stimuli. Supplementary Fig. 17b shows the effect of the time interval between the optical and electrical pulses on the conductance state of the BSO-EGT. As the interval time increases, the collected current value decreases significantly.

In order to evaluate the multimodal sensing capability of optoelectronic reservoirs, 4-bit binary streams of different fused input modes were applied. Classified into five combinations, namely “LLLL”, “EEEE”, “LLEL”, “LLEE”, and “LEEE”, respectively. As an example, mode “LLEE” denotes that the first two stimulation pulses are optical and the last two are electrical. As shown in Fig. 3a, each square wave input is considered as one bit, and the 4-bit input stream is encoded as a pulse sequence from “0000” to “1111”, in which the “off” and “on” state of the optical or electrical pulse denote “0” and “1”, respectively. Figure 3b shows the temporal variation of the drain currents for four different inputs, indicating that the final state of the reservoir is not only dependent on the last stimulus, but also related to the history of external stimuli. Figure 3c illustrates the evolution of the channel current after each pulse application for 16 combinations of inputs. Due to the nonlinear relaxation characteristics of the reservoir, its final state depends on its activity history. Therefore, “0001” and “0010” are two different sequences for the reservoir. When the input mode is “LLEL”, the final states can be distinguishable from each other even

with the same initial states (Fig. 3d). The distinguishability of the 16 states indicates the potential of our device for in-memory RC applications<sup>52</sup>. When only one type of external stimulus was applied (“LLLL” and “EEEE”), the BSO-EGT also showed good distinguishability between inputs with different coding sequences (Supplementary Fig. 18).

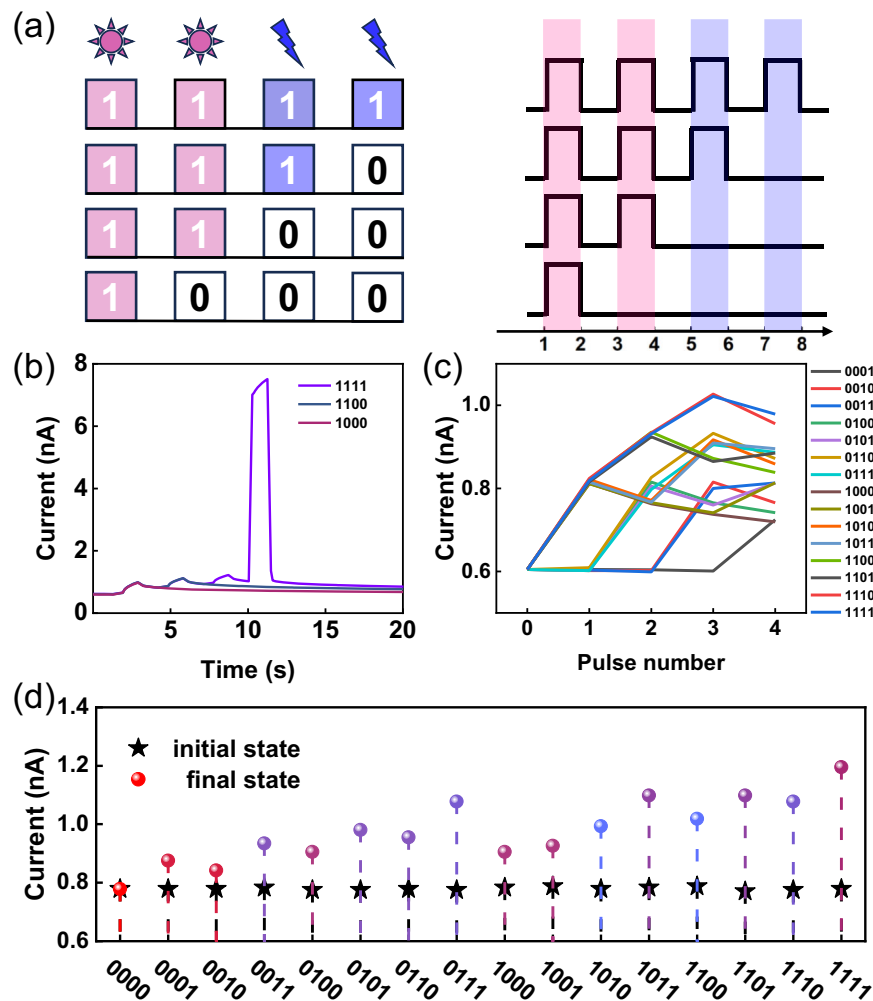
The distributions of 16-state values for different modes are distinguishable, so the classification tasks can ultimately be performed through computer simulation. The classification results for the other two multimodal inputs (“LLEE” and “LEEE”) were documented in Supplementary Fig. 19. It should be noted that as long as the relaxation process of the device has nonlinear and volatile characteristics, in theory, whether the potentiation or depression curves can be used to implement RC. Our as-prepared BSO-EGT is in a high-resistance state, the downward adjustment in the pristine state will cause significant leakage, making it difficult to perform RC based on the depression process. But there have been recent studies utilizing the depression behavior to implement reservoirs<sup>53,54</sup>.

### Static image recognition with reconfigurable BSO-EGT

Human perceive external information through a multi-sensory approach, while existing electronic systems use a relatively single method with a limited range of applications. A multi-sensory fusion method can be adopted to extend the application scope. Since the BSO transistors can respond to both optical and electrical inputs, it is possible to read contaminated (i.e., noisy or partially complete) image information using a hybrid optical-electrical approach. Furthermore, BSO-EGTs demonstrate both volatility and non-volatility in response to various external stimuli. The volatile nature, coupled with the non-linearity of the device, offers great potential as a reservoir for image pre-processing. In contrast, the non-volatile property fulfills the requirements of artificial synapses, as it facilitates post-processing and storage functionalities<sup>16,30</sup>. Based on the above-mentioned reconfigurable and optoelectronic sensing properties of BSO-EGT, we extracted the specific parameters of the device and further simulated a system integrating reservoir and ANN functions (Supplementary Fig. 20). Using this system, we first investigated the recognition of static polluted images.

Fashion-MNIST dataset were chosen for demonstration (see Supplementary Fig. 21 for image samples). For pictures of Fashion-MNIST partially contaminated by pigments (Fig. 4a), the left side can be perceived through optical signals, whereas the right side necessitates the assistance of pressure sensors to convert tactile sensations into electrical signals. Here, the contaminated part is defined as invisible information. The “pollution degree” or “invisible degree” indicates how many proportions of the pixels in the picture are contaminated. For ease of demonstration, we omit the sensing and processing steps of the piezoelectric sensors. The image information of the polluted part can no longer be obtained through optical perception, but can only be obtained through electrical signals. The detailed processing flow for contaminated images is shown in Fig. 4a. The original image (28 × 28 = 784 pixels) is reorganized into a stream of 196 × 4 pixels, after which the average value between pixels is used as the threshold to binarize the images. The processed image is read with two single-signal modes (“LLLL” and “EEEE”) and three mixed-signal modes (“LLEL”, “LLEE” and “LEEE”), respectively. The 196 groups of inputs are fed into the corresponding reservoir for pre-processing, which can compress the amount of data and map the low-dimensional feature space into a high-dimensional one using the nonlinearity process of the devices to facilitate the subsequent classification<sup>18,55</sup>.

An artificial neural network (ANN) is used for training and inference with the weight update model based on the experimental measurements. Figure 4b shows the outputs of the “LLLL” reservoirs with contamination degree of 10% (top panel) and 90% (bottom panel), respectively. Obviously, the network operating under the “LLLL” read



**Fig. 3 | Nonlinear mapping of multimodal signals based on EGT reservoirs.**

**a** Combination mode for 16 inputs of 4-bit optical and electrical pulses. Here, both the optical and electrical pulse widths were 1 s and the pulse interval was 1 s. **b** Response characteristics of the channel current in 1111, 1110, 1100, 1000

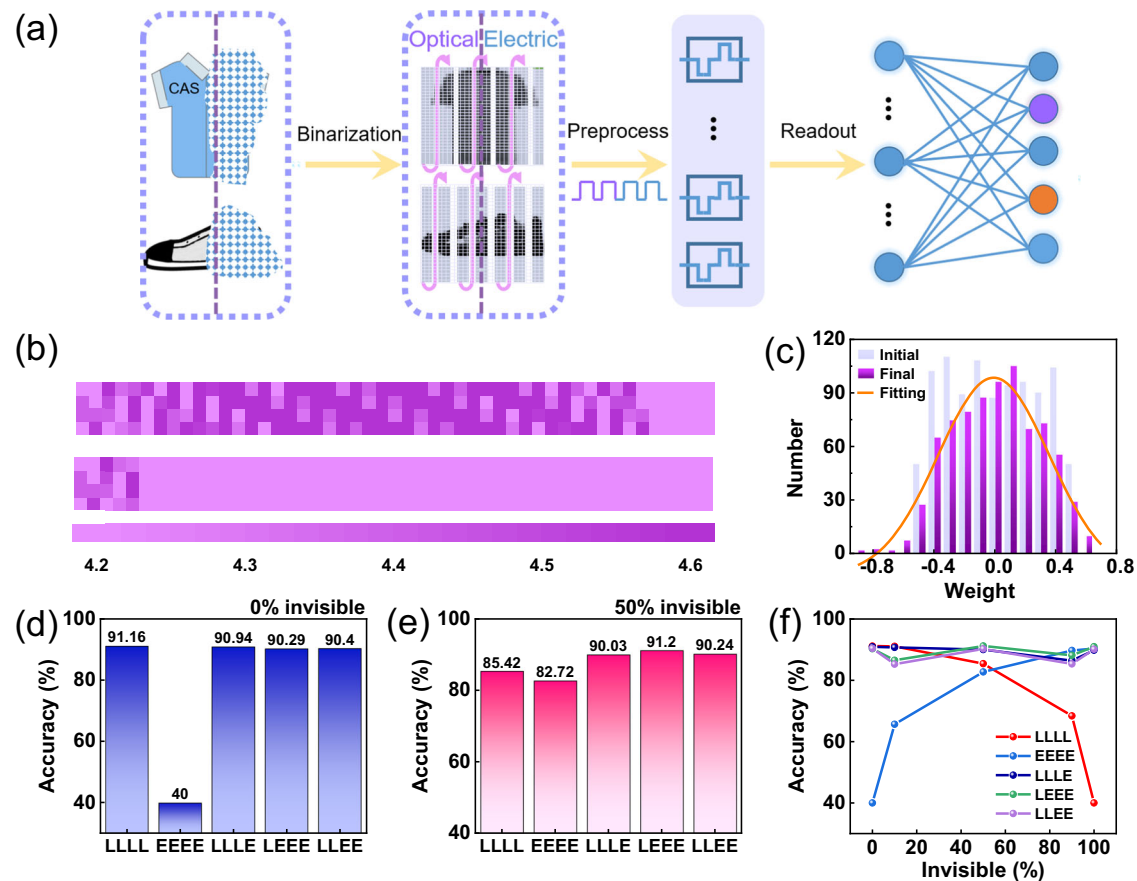
combinations under the “LLLE” mode input (light intensity  $70 \text{ mW/cm}^2$ ,  $V_G = 1 \text{ V}$ ). **c** The distinguishable output of 4-bit reservoir states reading at  $0.3 \text{ V}$ . **d** Initial and final current values of the reservoir state with pulse stimulation of 16 combinations in the “LLLE” inputs mode.

mode can obtain little effective information when there is a high percentage of contamination. Supplementary Fig. 22 gives the results of the other four read modes. Figure 4c illustrates the distribution of ANN synaptic weights before and after training. After the training process, the synaptic weights change from a random to a normal distribution, indicating that the neural network has been trained effectively. Figure 4d–f presents the final recognition accuracy obtained by the five read modes under different contamination levels. An invisibility of 0% means that the reservoirs of “LLLL” mode can acquire all the information from the image, while that of “EEEE” mode cannot obtain any content. Therefore, the “LLLL” mode is able to achieve a higher recognition rate as the classification is based on the actual information, while the inference of the “EEEE” model is indistinguishable from random guessing, so its recognition accuracy is much lower (Fig. 4d). And the situation is completely opposite in the case of 100% invisibility. The single-signal modes do not facilitate the effective image information acquisition in particularly extreme cases, while the mixed-signal modes allow for judgment and classification based on actual information due to its utilization of two reading channels. Thus, “LLLE”, “LLEE”, and “LEEE” can complete image recognition tasks with around 90% recognition accuracy at all contamination levels. When just one channel is used to obtain information, only parts of the features can be obtained, and the recognition accuracy is low. However,

the mixed mode can collect more comprehensive and rich information, so a higher recognition accuracy can be obtained. This result indicates that the mixed modes are more universal and can be applied to a wide range of complex situations.

### Multimodal dynamic gesture recognition with reconfigurable BSO-EGT

As a proof of concept, we mimic human perception of dynamic gesture recognition based on reconfigurable BSO-EGT. For gestures with spatiotemporal information, the decoupled sensation of vision and hearing may lead to misjudgment the direction or the object. If the two modes are combined for audio-visual fusion perception, the correct recognition accuracy can be greatly improved (Fig. 5a). Here, the EgoGesture dataset<sup>56,57</sup> was employed. Five gestures (Supplementary Fig. 23) were selected from eighty-three categories, and a sub-dataset with a sample size of 1250 was constructed. More details about dataset construction can be found in the “Methods”. Here, each sample contains four frames. Figure 5b represents the three-dimensional spatial map of the first one, in which the XY plane refers to the coordinates, while the Z-axis is the value of pixels. The color of the data points is a linear mapping of the pixel values, which enables the content of the picture to be analyzed based on the distribution of the color and position of the orbs. Supplementary Fig. 24 gives the spatial maps of



**Fig. 4 | Fused information input reservoir for recognition of multimodal Fashion-MNIST datasets.** **a** Image perception in complex environments. The right half of the information is invisible which can only be perceived through electrical signals, while the left half of information is visible which can only be perceived through visual signals. The original image ( $28 \times 28 = 764$  pixels) was reorganized into a  $196 \times 4$  by column vector array. **b** Output results of the reservoir in the “LLLL” sensing mode, when 10% (top) and 90% (bottom) of the image can only be read by

electrical signals. **c** The change of ANN synaptic weight before and after training. The trained synaptic weight conforms to the normal distribution. **d, e** Recognition accuracy under the five perception methods when 0% and 50% of the image can only be collected by electrical signals. **f** Recognition accuracy of the five information perception modes variation with the proportion of the contaminated part (which can only be read via electrical signals).

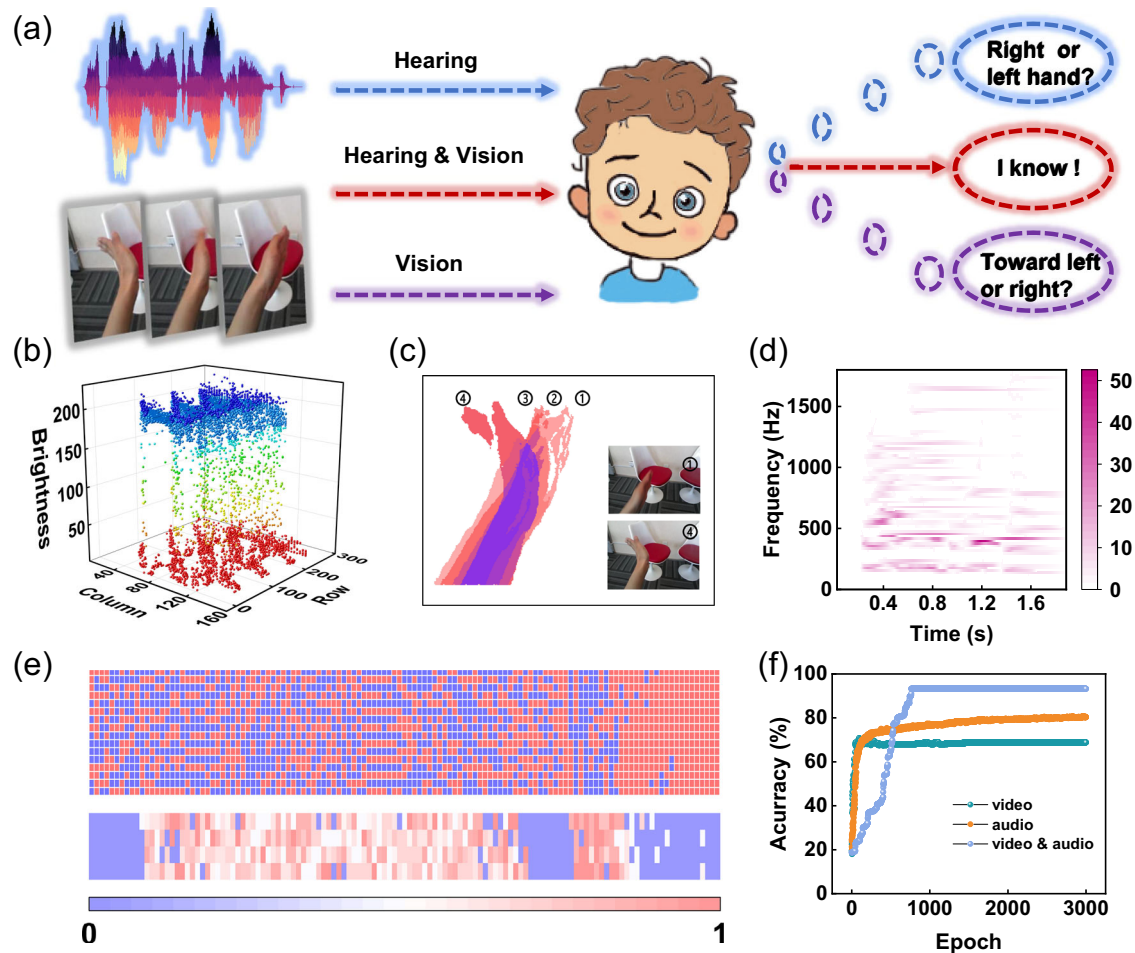
the other three frames of this sample. The four figures have large similarity, indicating that the contents of the four frames are similar. However, the color and spatial distributions of the data points differ in details, suggesting that they are images of different states of moving objects.

Similar to the static image processing process, the multi-frame pictures with temporal information can be mapped to the pixel matrix into a light pulse matrix, which is illuminated on the optical reservoir array to achieve information perception. More details can be found in “Methods”. The visualization diagram for the intermediate state shows the result of the superposition of the four frames after the pre-processing of the reservoirs (Fig. 5c). The last light pulse stimulation can be retained to the greatest extent, so the information of the last picture is clearer than that of the previous three. Furthermore, the overlapping part of the wrist in the four pictures will be stimulated with four consecutive pulses, so the output signal of the overlapping part was the strongest. Afterward, the pre-processed data would be fed into the neural network for subsequent training and inference. To avoid the slow training of processing units caused by simulation based on device parameters, the ANN in this part is built with the ideal model, while the reservoir is constructed based on device parameters.

Through 25 different public AI sound sources and text-to-speech conversion program, the names of the five selected actions are

converted into speech, and 125 original samples are generated. Then, these samples were Fourier-transformed to obtain the frequency domain information. Figure 5d shows the speech spectrogram of a sample, which reflects the frequency domain distribution of the audio at different time moments. The time-domain and frequency-domain information were sampled separately and combined into an input vector of length 2000. On this basis, 10% random noise is added to the audio input signals to simulate possible bit error ratio (BER) during digital signal transmission, and the number of samples in audio dataset was also expanded to 1250 to match the size of corresponding video dataset. More details about dataset construction can be found in the Methods sub-section “Dataset construction”. Afterward, the operations consistent with Fig. 4 were carried out, and Fig. 5e shows the input vector and the reservoir output.

After 3000 epochs, the accuracy of both video and audio recognition is lower than 80%, indicating that it is difficult to recognize different gestures with a single modal information (Fig. 5f). Since the BSO-EGT has the property of responding to multiple stimuli, a multi-sensory fusion neural network<sup>29,58</sup> using decision fusion method was constructed to demonstrate the multimodal gesture recognition. The audio and video datasets are processed through their respective reservoirs and neural networks, and then integrated and analyzed through the fusion-layer neural network. Evidently, the multimodal recognition method greatly improves the gesture classification



**Fig. 5 | Mimics human perception of dynamic audio-visual information.**

**a** Simplified diagram of humans achieving high levels of cognition through audio-visual integration. **b** Three-dimensional space mapping of “Rolling hand toward left” gesture, a sample from the EgoGesture dataset. **c** Images superimposed at four time points in the intermediate state after hand movements were preprocessed by the reservoir. **d** Speech spectrogram of “Rolling hand toward left” gesture,

reflecting the frequency domain distribution of audio at different moments. Corresponding voice overs for gestures, generated using text-to-speech engine. **e** Input vector after normalization of time and frequency domain information, and output vector after reservoir processing. **f** Comparison of recognition accuracy under single- and multi-mode information processing.

accuracy, which reached 94.24% after 3000 epochs (Fig. 5f). This result indicates that the multimodal recognition method based on reconfigurable BSO-EGT has superior performance. Compared with previous studies, our BSO-EGT has the advantages of precise regulation and comprehensive functions in terms of reconfigurability (Supplementary Table 1). At the application demonstration level, we verified its multi-sensory integration capabilities and realized advanced neuromorphic applications with the multimode sensing, storage, and processing capabilities of a single device.

#### Demonstration of wide application wavelength based on IGZO

In this work, we mainly propose a design principle for a kind of reconfigurable device. Channel materials include but are not limited to  $\text{BaSnO}_3$ . A variety of oxide materials with responses to light can serve as channels. For example,  $\text{InGaZnO}_4$  (IGZO), which is widely concerned by industry, can also be an option. IGZO has been mainly applied in industry due to its optical transparency, low processing temperature, and compatibility with various gate insulators<sup>59,60</sup>. It occurs the band-to-band excitation, the oxygen vacancies ionization, and the metastable peroxides formation in the a-IGZO semiconductor during the light illumination<sup>61,62</sup>. Due to the high density of the trap states in the gap, the high-quality a-IGZO films have a small conduction band tail ( $\sim 2.3$  eV)<sup>63</sup>, resulting in broad spectral responses. The use of

amorphous IGZO in thin film transistors offers numerous advantages, including excellent uniformity, high mobility, high switching current ratio, and large-scale processing<sup>64</sup>.

We grew 15 nm a-IGZO film on  $\text{SiO}_2/\text{Si}$  substrates and X-ray photoelectron spectroscopy (XPS) results showed clear characteristic peaks of  $\text{In } 3d$ ,  $\text{Ga } 2p$ , and  $\text{Zn } 2p$ , respectively, indicating that the IGZO film is of high quality (Supplementary Fig. 25). IGZO was used as the channel material to fabricate an electrolyte-gated transistor (IGZO-EGT) with the same structure. Based on the same experiment scheme, IGZO showed the potential to achieve the same functionalities we demonstrated with BSO film. IGZO-EGT also demonstrates the ability to accurately switch between volatile and non-volatile modes affected by voltage amplitude (Supplementary Fig. 26). The response of IGZO-EGT at green (532 nm), blue (450 nm), and ultraviolet (375 nm) light wavelengths was also tested (Supplementary Fig. 27) and showed the wide applicable light wavelengths of IGZO-EGT. Moreover, a clear nonlinear relaxation process was observed in IGZO-EGT after the electrical or optical stimulation was removed, indicating its potential to realize reservoirs. Therefore, using the design concept we proposed, electrolyte transistors using IGZO as the channel material can also achieve reconfigurable reservoir and artificial synapse functions, and then complete complex tasks such as audio-visual fusion recognition.



## Discussion

We have reported a neuromorphic EGT with multimode sensing, memory, and processing capabilities. The devices exhibit dynamical processes with tunable time scales under optical and electrical modulation, and are capable of realizing reconfigurable functions between physical reservoir and artificial synapse. A parallel optoelectronic fusion system composed of reservoir and ANN functions was simulated based on reconfigurable BSO-EGT. The Fashion-MNIST dataset containing multiple types of information was utilized as a standard test, and the recognition accuracy above 90% shows the superiority of this system in information processing. Furthermore, dynamic gesture recognition was also used to test the system performance. The higher recognition accuracy under audio-visual integration demonstrates the advantages of our constructed system in simulating biological multisensory fusion. Moreover, the approach demonstrated in our study can be utilized to a broad range of materials, as long as its relaxation process under stimuli has non-linear and volatile characteristics. Benefiting from the multi-modal sensing, storage and processing performance of our proposed device, complete audio-visual integration and recognition tasks can be realized on a single device. The processing of dynamic tasks reflects the real-time processing capability of our proposed device. The proposed system could advance the development of a multi-sensory human-machine interaction platform.

## Methods

### Sample preparation

The 10 nm BaSnO<sub>3</sub> film was epitaxially grown on (001)-oriented MgO substrates at 780 °C under O<sub>2</sub> pressure of 5.5 Pa. Pulsed laser deposition was used with a 308-nm XeCl excimer laser, with an energy density of about 1 J/cm<sup>2</sup> and a repetition of 3 Hz. The samples were cooled down to room temperature at 20 °C/min. The growth conditions were optimized to minimize the cation non-stoichiometry induced defects of as-grown BaSnO<sub>3</sub> films by adjusting the stoichiometric accuracy ([Sn]/[Ba] = 1) through a suitable target-to-substrate distance ( $d_{ts}$  = 50 mm).

The 15 nm a-IGZO film was grown on SiO<sub>2</sub>/Si substrates at room temperature under O<sub>2</sub> pressure of 5 Pa through pulsed laser deposition. Afterward, the film was annealed at 300 °C for 1 h. Pulsed laser deposition was equipped with a 308-nm XeCl excimer laser, with an energy density of about 1 J/cm<sup>2</sup> and a repetition of 3 Hz. The samples were cooled down to room temperature at 20 °C/min.

### Device fabrication

Through photolithography and ion beam etching technology, the BSO film is patterned into channels with an effective area of 50 × 180 μm<sup>2</sup>. A coplanar side-gate structure is adopted, and a 30 nm Pt layer is deposited as an electrode by magnetron sputtering. The distance between the gate and the channel is 10 μm. The transistor device was completed by dropping an ionic liquid N, N-diethyl-n-(2-methoxyethyl)-N-methylammoniumbis-(trifluoromethylsulphonyl)-imide (DEME-TFSI) on the channel and gate electrodes.

### Material characterization

X-ray diffraction patterns of the BaSnO<sub>3</sub> film was performed using a Rigaku SmartLab instrument with a 2θ range from 35 to 50° in step of 0.05°. STEM imaging was conducted by a double Cs-corrected JEOL JEM-ARM200CF operated at 200 kV with a CEOS Cs corrector (CEOS GmbH, Heidelberg, Germany). HAADF-STEM images were recorded with collection semi-angles of 90–370 mrad. Optical transmittance spectra were taken in air at room temperature with spectrophotometers (Cary 5000 UV-Vis-NIR, Agilent and Excalibur3100, Varian). In order to reveal the relationship between hydrogen concentration and gating voltages, a TOF-SIMS system (ION-TOF GmbH)

was used to identify the depth profiles of protons. XPS measurements were performed on ThermoFisher Scientific ESCALAB 250X under monochromatic Al Kα radiation with an energy of 1486.6 eV.

### Device characterization

All the electrical characterizations were measured in a Laskeshore probe station with a Keithley 4200 semiconductor parameter analyzer in vacuum at room temperature. An UV laser at a wavelength of 375 nm and a blue laser at 450 nm were used for optical excitation in the experiments.

### Dataset construction

The video classification task uses the public EgoGesture dataset, which contains a total of 83 different gesture. Each category contains a large number of samples and each sample in this dataset contains dozens of chronologically arranged frames describing the corresponding gesture. We selected 5 gestures from 83 categories and the first 250 samples from each corresponding category are retained to construct a sub-dataset with a sample size of 1250.

The audio dataset is generated based on the built video sub-dataset. Through 25 different public AI sound sources, the names of the five selected gestures from EgoGesture dataset are converted into speeches, and 125 original samples are generated. After pre-processing and adding noise, the volume of the audio dataset also expands to 1250. In the video and audio datasets, the ratio of training set and test set is 1:1 which means each category contains 625 samples.

### Video classification

Each sample in the EgoGesture dataset contains dozens of frames. To compress the amount of data, we randomly extracted four frames in chronological order. The size of each frame is 320 pixels × 240 pixels, which can be represented by a binary matrix of size [320, 240] after binarization. Then, we expanded the matrix to get a one-dimensional column vector with the size of [76,800, 1]. After merging the vectors corresponding to the four frames, every final input needs to be represented by a matrix of size [76,800, 4]. Then, each row of the matrix was encoded with light pulses and fed into the corresponding reservoir. After being processed and compressed by the reservoirs, the input of the ANN is a vector of length 76,800. Therefore, a two-layer fully connected neural network with a size of [76,800, 5000, 5] was used to perform the video classification recognition task. The ReLU function and the cross-entropy loss function were selected as the activation functions of the hidden layer and readout layer respectively, and the weights were updated based on the back propagation algorithm. The audio classification used the same approach, while the difference is that the size of the network. This task only required a single-layer neural network, and the activation function of the output layer is also a cross-entropy loss function.

### Data availability

Source data for the figures are provided as a Source data file. All relevant data within the Supplementary Information are available from the corresponding authors upon request. Source data are provided with this paper.

### Code availability

All code used in simulations supporting this article is available from the corresponding authors upon request.

### References

1. Solvi, C., Gutierrez Al-Khudhairy, S. & Chittka, L. Bumble bees display cross-modal object recognition between visual and tactile senses. *Science* **367**, 910–912 (2020).

2. Park, H. L. et al. Flexible neuromorphic electronics for computing, soft robotics, and neuroprosthetics. *Adv. Mater.* **32**, e1903558 (2020).
3. Jiang, C. et al. Mammalian-brain-inspired neuromorphic motion-cognition nerve achieves cross-modal perceptual enhancement. *Nat. Commun.* **14**, 1344 (2023).
4. Ban, Y., Alameda-Pineda, X., Girin, L. & Horaud, R. Variational Bayesian inference for audio-visual tracking of multiple speakers. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 1761–1776 (2021).
5. Qian, X., Madhavi, M., Pan, Z., Wang, J. & Li, H. Multi-target DoA estimation with an audio-visual fusion mechanism. In *ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2021).
6. Li, G. et al. Photo-induced non-volatile VO<sub>2</sub> phase transition for neuromorphic ultraviolet sensors. *Nat. Commun.* **13**, 1729 (2022).
7. Ahmed, T. et al. Fully light-controlled memory and neuromorphic computation in layered black phosphorus. *Adv. Mater.* **33**, e2004207 (2021).
8. Moin, A. et al. A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition. *Nat. Electron.* **4**, 54–63 (2021).
9. Feldmann, J., Youngblood, N., Wright, C. D., Bhaskaran, H. & Pernice, W. H. P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
10. Ji, X. et al. Mimicking associative learning using an ion-trapping non-volatile synaptic organic electrochemical transistor. *Nat. Commun.* **12**, 2480 (2021).
11. Torricelli, F. et al. Electrolyte-gated transistors for enhanced performance bioelectronics. *Nat. Rev. Methods Prim.* **1**, 66 (2021).
12. Lanza, M. et al. Memristive technologies for data storage, computation, encryption, and radio-frequency communication. *Science* **376**, eabj9979 (2022).
13. Gan, C., Zhang, Y., Wu, J., Gong, B. & Tenenbaum, J. B. Look, listen, and act: towards audio-visual embodied navigation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)* (2020).
14. Qian, X., Wang, Z., Wang, J., Guan, G. & Li, H. Audio-visual cross-attention network for robotic speaker tracking. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **31**, 550–562 (2023).
15. Keshavarzi, S. et al. Multisensory coding of angular head velocity in the retrosplenial cortex. *Neuron* **110**, 532–543.e539 (2022).
16. Zhang, Z. et al. In-sensor reservoir computing system for latent fingerprint recognition with deep ultraviolet photo-synapses and memristor array. *Nat. Commun.* **13**, 6590 (2022).
17. Tan, H. & van Dijken, S. Dynamic machine vision with retinomorphic photomemristor-reservoir computing. *Nat. Commun.* **14**, 2169 (2023).
18. Appeltant, L. et al. Information processing using a single dynamical node as complex system. *Nat. Commun.* **2**, 468 (2011).
19. Milano, G. et al. In materia reservoir computing with a fully memristive architecture based on self-organizing nanowire networks. *Nat. Mater.* **21**, 195–202 (2022).
20. Liu, K. et al. Multilayer reservoir computing based on ferroelectric  $\alpha$ -In<sub>2</sub>Se<sub>3</sub> for hierarchical information processing. *Adv. Mater.* **34**, e2108826 (2022).
21. Wakabayashi, S., Arie, T., Akita, S., Nakajima, K. & Takeji, K. A multi-tasking flexible sensor via reservoir computing. *Adv. Mater.* **34**, e2201663 (2022).
22. Sun, L. et al. In-sensor reservoir computing for language learning via two-dimensional memristors. *Sci. Adv.* **7**, eabg1455 (2021).
23. Jeong, B., Gkoupidenis, P. & Asadi, K. Solution-processed perovskite field-effect transistor artificial synapses. *Adv. Mater.* **33**, e2104034 (2021).
24. Kumar, D., Li, H., Das, U. K., Syed, A. M. & El-Atab, N. Flexible solution-processable black-phosphorus-based optoelectronic memristive synapses for neuromorphic computing and artificial visual perception applications. *Adv. Mater.* **35**, e2300446 (2023).
25. Wang, W. et al. A memristive deep belief neural network based on silicon synapses. *Nat. Electron.* **5**, 870–880 (2022).
26. Liang, X., Luo, Y., Pei, Y., Wang, M. & Liu, C. Multimode transistors and neural networks based on ion-dynamic capacitance. *Nat. Electron.* **5**, 859–869 (2022).
27. Wang, S. et al. An organic electrochemical transistor for multimodal sensing, memory and processing. *Nat. Electron.* **6**, 281–291 (2023).
28. Wan, T. et al. In-sensor computing: materials, devices, and integration technologies. *Adv. Mater.* **1**, e2203830 (2022).
29. Wang, M. et al. Gesture recognition using a bioinspired learning architecture that integrates visual data with somatosensory data from stretchable sensors. *Nat. Electron.* **3**, 563–570 (2020).
30. Liu, K. et al. An optoelectronic synapse based on  $\alpha$ -In<sub>2</sub>Se<sub>3</sub> with controllable temporal dynamics for multimode and multiscale reservoir computing. *Nat. Electron.* **5**, 761–773 (2022).
31. Liu, M. et al. A star-nose-like tactile-olfactory bionic sensing array for robust object recognition in non-visual environments. *Nat. Commun.* **13**, 79 (2022).
32. John, R. A. et al. Reconfigurable halide perovskite nanocrystal memristors for neuromorphic computing. *Nat. Commun.* **13**, 2074 (2022).
33. Ge, C. et al. Gating-induced reversible H<sub>x</sub>VO<sub>2</sub> phase transformations for neuromorphic computing. *Nano Energy* **67**, 104268 (2020).
34. Liu, X. et al. Near-sensor reservoir computing for gait recognition via a multi-gate electrolyte-gated transistor. *Adv. Sci.* **10**, e2300471 (2023).
35. Yang, J. T. et al. Artificial synapses emulated by an electrolyte-gated tungsten-oxide transistor. *Adv. Mater.* **30**, e1801548 (2018).
36. Chambers, S. A., Kaspar, T. C., Prakash, A., Haugstad, G. & Jalan, B. Band alignment at epitaxial BaSnO<sub>3</sub>/SrTiO<sub>3</sub>(001) and BaSnO<sub>3</sub>/LaAlO<sub>3</sub>(001) heterojunctions. *Appl. Phys. Lett.* **108**, 152104 (2016).
37. Raghavan, S. et al. High-mobility BaSnO<sub>3</sub> grown by oxide molecular beam epitaxy. *Appl. Mater.* **4**, 016106 (2016).
38. Prakash, A. et al. Wide bandgap BaSnO<sub>3</sub> films with room temperature conductivity exceeding 10<sup>4</sup> S cm<sup>-1</sup>. *Nat. Commun.* **8**, 15167 (2017).
39. Lee, Y. et al. Reversible manipulation of photoconductivity caused by surface oxygen vacancies in perovskite stannates with ultraviolet light. *Adv. Mater.* **34**, e2107650 (2022).
40. Yun, H. et al. Metallic line defect in wide-bandgap transparent perovskite BaSnO<sub>3</sub>. *Sci. Adv.* **7**, eabd4449 (2021).
41. Ganguly, K. et al. Structure and transport in high pressure oxygen sputter-deposited BaSnO<sub>3- $\delta$</sub> . *Appl. Mater.* **3**, 062509 (2015).
42. Park, J., Kim, U. & Char, K. Photoconductivity of transparent perovskite semiconductor BaSnO<sub>3</sub> and SrTiO<sub>3</sub> epitaxial thin films. *Appl. Phys. Lett.* **108**, 092106 (2016).
43. Wang, S. et al. A MoS<sub>2</sub>/PTCDA modulation heterojunction synapse with efficient photoelectric dual modulation and versatility. *Adv. Mater.* **31**, 1806227 (2019).
44. Meng Y, et al. Artificial visual systems enabled by quasi-two-dimensional electron gases in oxide superlattice nanowires. *Sci. Adv.* **6**, eabc6389 (2020).
45. Taylor, G. W. & Simmons, J. G. Basic equations for statistics, recombination processes, and photoconductivity in amorphous insulators and semiconductors. *J. Non-Cryst. Solids* **8–10**, 940–946 (1972).
46. De Souza, R. A. Oxygen diffusion in SrTiO<sub>3</sub> and related perovskite oxides. *Adv. Funct. Mater.* **25**, 6326–6342 (2015).
47. Lee, W.-J. et al. Oxygen diffusion process in a Ba<sub>0.96</sub>La<sub>0.04</sub>SnO<sub>3</sub> thin film on SrTiO<sub>3</sub>(001) substrate as investigated by time-dependent

- Hall effect measurements. *Phys. Status Solidi (A)* **212**, 1487–1493 (2015).
48. Ge, C. et al. Metal-insulator transition induced by oxygen vacancies from electrochemical reaction in ionic liquid-gated manganite films. *Adv. Mater. Interfaces* **2**, 1500407 (2015).
49. Bisri, S. Z., Shimizu, S., Nakano, M. & Iwasa, Y. Endeavor of iontronics: from fundamentals to applications of ion-controlled electronics. *Adv. Mater.* **29**, 1607054 (2017).
50. Lu, N. et al. Electric-field control of tri-state phase transformation with a selective dual-ion switch. *Nature* **546**, 124–128 (2017).
51. Yuan, H. T. et al. Hydrogenation-induced surface polarity recognition and proton memory behavior at protic-ionic-liquid/oxide electric-double-layer interfaces. *J. Am. Chem. Soc.* **132**, 6672–6678 (2010).
52. Du, C. et al. Reservoir computing using dynamic memristors for temporal information processing. *Nat. Commun.* **8**, 2204 (2017).
53. Chen, Z. et al. All-ferroelectric implementation of reservoir computing. *Nat. Commun.* **14**, 3585 (2023).
54. Liu, Z. et al. Interface-type tunable oxygen ion dynamics for physical reservoir computing. *Nat. Commun.* **14**, 7176 (2023).
55. Wu, X. S. et al. Wearable in-sensor reservoir computing using optoelectronic polymers with through-space charge-transport characteristics for multi-task learning. *Nat. Commun.* **14**, 468 (2023).
56. Zhang, Y., Cao, C., Cheng, J. & Lu, H. EgoGesture: a new dataset and benchmark for egocentric hand gesture recognition. *IEEE Trans. Multimed.* **20**, 1038–1050 (2018).
57. Cao, C., Zhang, Y., Wu, Y., Lu, H. & Cheng, J. Egocentric gesture recognition using recurrent 3D convolutional neural networks with spatiotemporal transformer modules. In *2017 IEEE International Conference on Computer Vision (ICCV)* (2017).
58. Tuia, D., Volpi, M. & Moser, G. Decision fusion with multiple spatial supports by conditional random fields. *IEEE Trans. Geosci. Remote Sens.* **56**, 3277–3289 (2018).
59. Hays, D. C., Gila, B. P., Pearton, S. J. & Ren, F. Energy band offsets of dielectrics on InGaZnO<sub>4</sub>. *Appl. Phys. Rev.* **4**, 021301 (2017).
60. Jang, Y., Park, J., Kang, J. & Lee, S.-Y. Amorphous InGaZnO (a-IGZO) synaptic transistor for neuromorphic computing. *ACS Appl. Electron. Mater.* **4**, 1427–1448 (2022).
61. Ke, S. et al. Indium-gallium-zinc-oxide based photoelectric neuromorphic transistors for modulable photoexcited corneal nociceptor emulation. *Adv. Electron. Mater.* **7**, 2100487 (2021).
62. Li, H. K. et al. A light-stimulated synaptic transistor with synaptic plasticity and memory functions based on InGaZnO<sub>x</sub>-Al<sub>2</sub>O<sub>3</sub> thin film structure. *J. Appl. Phys.* **119**, 244505 (2016).
63. Ide, K. et al. Effects of excess oxygen on operation characteristics of amorphous In-Ga-Zn-O thin-film transistors. *Appl. Phys. Lett.* **99**, 093507 (2011).
64. Liu, H. et al. High performance and hysteresis-free a-IGZO thin film transistors based on spin-coated hafnium oxide gate dielectrics. *IEEE Electron Device Lett.* **44**, 1508–1511 (2023).

## Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2019YFA0308500 to K.J.), the National Natural Science Foundation of China (No. 12222414 to C.G., No. 12074416 to C.G., No. 11721404 to K.J., No. 12174437 to C.W., No. 62075142 to Q.L.Z.), and the Youth Innovation Promotion Association of CAS (No. Y2022003 to C.G.).

## Author contributions

C.G. initiated the research. C.G. and K.J. supervised the project. P.L., G.L., and Z.L. prepared the sample. P.L., M.Z. and D.X. fabricated the device. The device measurements were done by P.L. with support from Z.W. Q.H.Z., and L.G. contributed to STEM measurements. Simulations were performed by M.Z. and D.X. P.L., M.Z., and C.G. wrote the manuscript. Q.L.Z., E.G., M.H., C.W., K.J., and G.Y. participated in the discussion of manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-47580-2>.

**Correspondence** and requests for materials should be addressed to Kuijuan Jin or Chen Ge.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

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

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

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

© The Author(s) 2024