

Fully integrated multi-mode optoelectronic memristor array for diversified in-sensor computing

Received: 23 May 2023

Accepted: 26 August 2024

Published online: 8 November 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

In-sensor computing, which integrates sensing, memory and processing functions, has shown substantial potential in artificial vision systems. However, large-scale monolithic integration of in-sensor computing based on emerging devices with complementary metal–oxide–semiconductor (CMOS) circuits remains challenging, lacking functional demonstrations at the hardware level. Here we report a fully integrated 1-kb array with 128×8 one-transistor one-optoelectronic memristor (OEM) cells and silicon CMOS circuits, which features configurable multi-mode functionality encompassing three different modes of electronic memristor, dynamic OEM and non-volatile OEM (NV-OEM). These modes are configured by modulating the charge density within the oxygen vacancies via synergistic optical and electrical operations, as confirmed by differential phase-contrast scanning transmission electron microscopy. Using this OEM system, three visual processing tasks are demonstrated: image sensory pre-processing with a recognition accuracy enhanced from 85.7% to 96.1% by the NV-OEM mode, more advanced object tracking with 96.1% accuracy using both dynamic OEM and NV-OEM modes and human motion recognition with a fully OEM-based in-sensor reservoir computing system achieving 91.2% accuracy. A system-level benchmark further shows that it consumes over 20 times less energy than graphics processing units. By monolithically integrating the multi-functional OEMs with Si CMOS, this work provides a cost-effective platform for diverse in-sensor computing applications.

Vision is the most important sense for human perception and understanding, as the brain acquires about 80% of its information through visual processing^{1,2}. Recent studies have suggested that the human visual system can be considered a hierarchical processor, from low-level pre-processing to high-level classification and cognition³. Low-level sensory processing extracts vital information from the massive amount of data collected from the environment. This information is then abstracted and refined through high-level processing units⁴. Low-level processing tasks include filtering⁵, noise reduction (or denoising)⁶ and feature enhancement⁷, whereas high-level sensory tasks, such as

recognition⁸ and tracking⁹, require multi-mode computation. Thus, integrating various sensor and computing units to achieve different levels of processing on a single substrate is highly desired but also challenging, for which in-sensor computing emerges as a promising solution. This paradigm could notably improve the processing efficiency and simplify the hardware system^{10–12}.

To accomplish different levels of in-sensor computing, various artificial neural network machine learning algorithms are essential for implementation^{13,14}, for example, convolutional neural networks (CNNs)^{15–17} and reservoir computing (RC)^{18–20}. However, different neural

✉ e-mail: wangyuyan@tsinghua.edu.cn; jtang@tsinghua.edu.cn; wuhq@tsinghua.edu.cn

networks usually require different physical elements to implement. For instance, in-sensor RC is a promising neuromorphic vision information processing paradigm²¹. It is advantageous in training (as only the read-out layer needs to be trained) and exhibits high efficiency and low latency^{22,23}. The RC system generally comprises a reservoir layer with dynamic nonlinear nodes and a non-volatile read-out layer that uses a weighted sum module^{13,21}. Owing to the different characteristics of these two layers, fully integrating an RC system in one substrate is challenging. Therefore, finding novel devices capable of working in both dynamic and non-volatile switching modes is of great interest to the field of in-sensor computing.

Optoelectronic memristors (OEMs) are promising candidates for such in-sensor computing, exhibiting resistive switching properties that can be modulated by both optical and electrical stimuli^{24–26}. Material wise, ZnO represents a favourable candidate for fabricating OEMs owing to its reliable electrical resistive switching characteristic^{27–29}, configurable photoelectric response^{30,31} and good compatibility with the silicon complementary metal–oxide–semiconductor (CMOS) backend-of-the-line integration process³². Recent studies have demonstrated that applying an electric field can alter the zinc interstitials and oxygen vacancies within the ZnO layer^{27,33}, leading to modifications in the defect levels and thus modulating the photoelectric response.

In this Article, we report a 128×8 crossbar array of one-transistor one-OEM (1T1OEM) cells integrated with Si CMOS circuits. The OEMs with a material stack of Pd/TiO_x/ZnO/TiN exhibit multi-modal characteristics by modulating the charge density distribution through synergistic electrical and optical operations. The underlying mechanism is revealed by the redistribution of the electric field density within the ZnO using differential phase-contrast scanning transmission electron microscopy (DPC-STEM). Furthermore, we demonstrate three representative visual processing applications using this multi-modal OEM system, ranging from low-level pre-processing simulations for noise reduction to high-level hardware experiments of human motion recognition. The OEM system features a CMOS-compatible fabrication process and configurable multi-mode functionality to support various neural networks (such as RC and CNN), providing a promising platform for multi-scenario signal processing.

Multi-functional 1T1OEM array

Here, we fabricate a 128×8 crossbar array on the basis of 1T1OEM cells on top of Si CMOS circuits (including decoders and drivers; Supplementary Fig. 1). Figure 1a shows an optical micrograph of the crossbar array and cross-sectional high-resolution transmission electron microscopy (TEM) image of the 1T1OEM cell. The detailed fabrication process is described in Methods. Here, the Ti layer scavenges oxygen from the ZnO layer, leading to its oxidation into TiO_x. This oxidation process is confirmed by the energy-dispersive X-ray spectroscopy (EDS) (Fig. 1a, right) and X-ray photoelectric spectroscopy (Supplementary Fig. 2) results. The light absorption spectrum (Supplementary Fig. 3) reveals that both ZnO and TiO_x exhibit high absorption rates in the ultraviolet to the visible region. This can be attributed to the presence of oxygen vacancies and other defects.

As illustrated in Fig. 1b, the OEM chip circuit is designed to showcase its capability for versatile in-sensor computing. Moreover, the OEM exhibits multi-mode resistive switching behaviours. Through the electroforming process, which generates oxygen vacancy conductive filaments, the dynamic OEM (D-OEM) mode can be transformed into the non-volatile OEM (NV-OEM) mode. As a result, these different OEM modes can be integrated on one wafer to simulate the human visual system. Implementing these multi-functional capabilities within the same array substantially reduces manufacturing costs, minimizes the hardware's physical size and eliminates the need for complex inter-chip information transfer, as shown in Fig. 1c. In addition, it can be customized to support various application scenarios and algorithms, such as CNN and RC. To demonstrate the versatility of our multi-mode OEM

array, we present diversified in-sensor computing tasks, including low-level sensory processing and high-level sensory processing as shown in Fig. 1d.

Figure 2a shows the pristine device structure and symmetric distribution of the Zn–O bond's atomic electric field from DPC-STEM. As depicted in Fig. 2b, the dynamic photocurrent behaviour exhibits short-term plasticity as observed in biological synapses, and consecutive light pulses with varying intensities can modulate it. The OEM operates in the D-OEM in the initial state, which can be attributed to the stimulated emission and rapid recombination of shallow interfacial defect states. Supplementary Fig. 4a–e illustrates the band structure change process during dynamic response. Figure 2c shows the paired-pulse facilitation (PPF) characteristics. The inset shows the correlation between the PPF index (A_2/A_1) and the interval time, which can be fitted with a double-exponential function

$$\text{PPF index} = 1 + C_1 e^{(-\Delta t/\tau_1)} + C_2 e^{(-\Delta t/\tau_2)} \quad (1)$$

where $\tau_1 = 0.3$ ms and $\tau_2 = 0.34$ s are the time constant of the rapid and slow relaxation phases, respectively. The results indicate that the ratio increases when the interval between the two pulses decreases.

The excitatory post-synaptic current (EPSC) is also affected by both the pulse width and the intensity of the light stimulus. As demonstrated in Fig. 2d, the amplitude of EPSC increases with wider pulse width. The EPSC response is measured down to the minimum pulse width of 10 μ s, as depicted in Supplementary Fig. 5, which is the limitation of the equipment used. As shown in Supplementary Fig. 6, the EPSC is also enhanced by increasing the light intensity and the duration of the pulse width. Furthermore, increasing the light pulse frequency from 5 Hz to 10 Hz to 50 Hz enhances the amplitude of EPSC (Supplementary Fig. 7). These short-term plasticity behaviours of D-OEM serve as a foundation for signal encoding and RC implementation using this mode^{8,19}.

Figure 2e illustrates the electrical forming process of OEM. In contrast to the pristine state, ZnO exhibits clear filament structures after the forming process. Meanwhile, as evidenced by DPC-STEM (Supplementary Fig. 8a,b), it reveals a transformation of the Zn–O bond's atomic electric field from a symmetric to an asymmetric distribution. It shows an enhancement of electric field intensity on one side of the Zn–O bond. This proves that the formation of oxygen vacancies (V_o , V_o^+ and V_o^{++}) leads to an inhomogeneous distribution of charge density around the oxygen atoms. These oxygen vacancy defect states (V_o^+ and V_o^{++}) provide a metallic conduction channel or hopping pathway for electron transport, increasing the device conductance, as shown in Supplementary Fig. 4f–j. Figure 2f illustrates the I – V characteristics of the OEM under electronic memristor (EM) mode. To evaluate the device's stability^{34,35}, Fig. 2g shows the cycling endurance between low-resistance state (LRS) and high-resistance state (HRS) for 10^9 cycles, and the schematic diagram of the test waveform is shown in Supplementary Fig. 9. We assess the retention of the EM by reading eight representative resistance states throughout 5,000 s at 85 °C, as depicted in Fig. 2h. Using the extracted activation energy, the extrapolated retention time at room temperature is 11.1 and 7.2 years for LRS and HRS, respectively (Supplementary Fig. 10). Meanwhile, the EM mode exhibits excellent multi-state analogue-resistive switching characteristics (Supplementary Fig. 11). During the forming process, the EDS results show that the oxygen ion content in the conductive filaments is lower than elsewhere (Supplementary Fig. 12). This phenomenon also confirms that more oxygen vacancies and trapping defects are generated in the OEM after the electroforming process.

The memristor operates in the NV-OEM mode after the forming process, as shown in Fig. 2i. Meanwhile, moiré fringes in the formed conductive filaments and an extremely enhanced electric field strength near the position of the oxygen atom are observed through DPC-STEM (Supplementary Fig. 8c). This phenomenon can be understood as phase

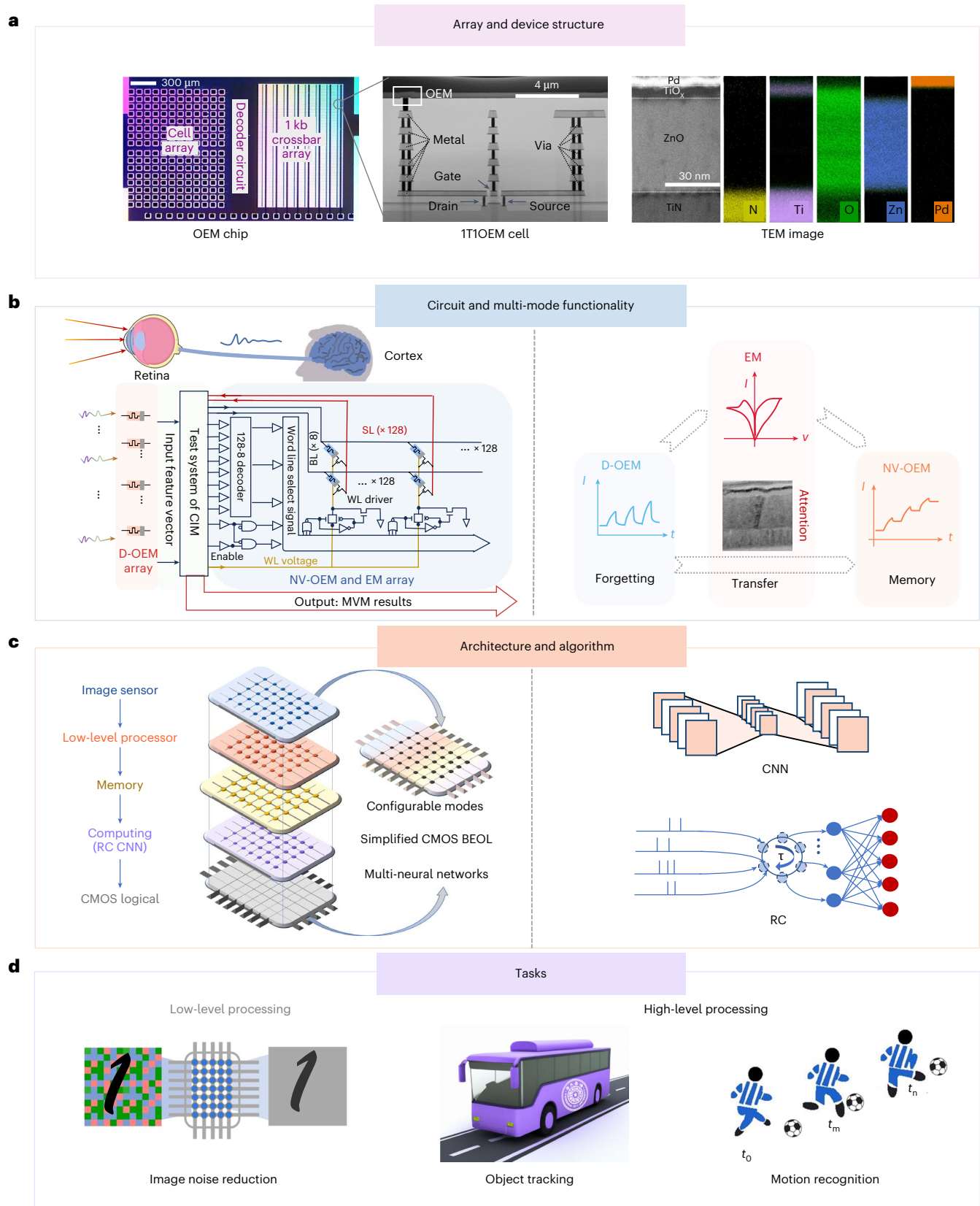


Fig. 1 | OEM array and in-sensor computing system. a, An optical micrograph of the 128×8 crossbar array (left), a cross-sectional high-resolution TEM image showing the 1T1OEM structure with a Pd/TiO₂/ZnO/TiN stack of $700 \times 700 \text{ nm}^2$ on top of the drain terminal of the Si transistor (centre) and EDS mapping of the OEM device showing the elemental distribution (right). **b**, A schematic diagram of the OEM chip circuit design for in-sensor computing and the configurable multi-functional OEM device illustrating the device current (I) versus time (t)

and applied voltage (V). CIM, computing-in-memory; WL, word line; SL, source line; BL, bit line; EM, electronic memristor. **c**, An illustration of the all-in-one architecture of the OEM system on one wafer and an illustration of the CNN and RC algorithm. BEOL, backend-of-the-line. **d**, An illustration of multi-scenario demonstrations from low-level image noise reduction processing to high-level object tracking and motion recognition processing using the OEM system.

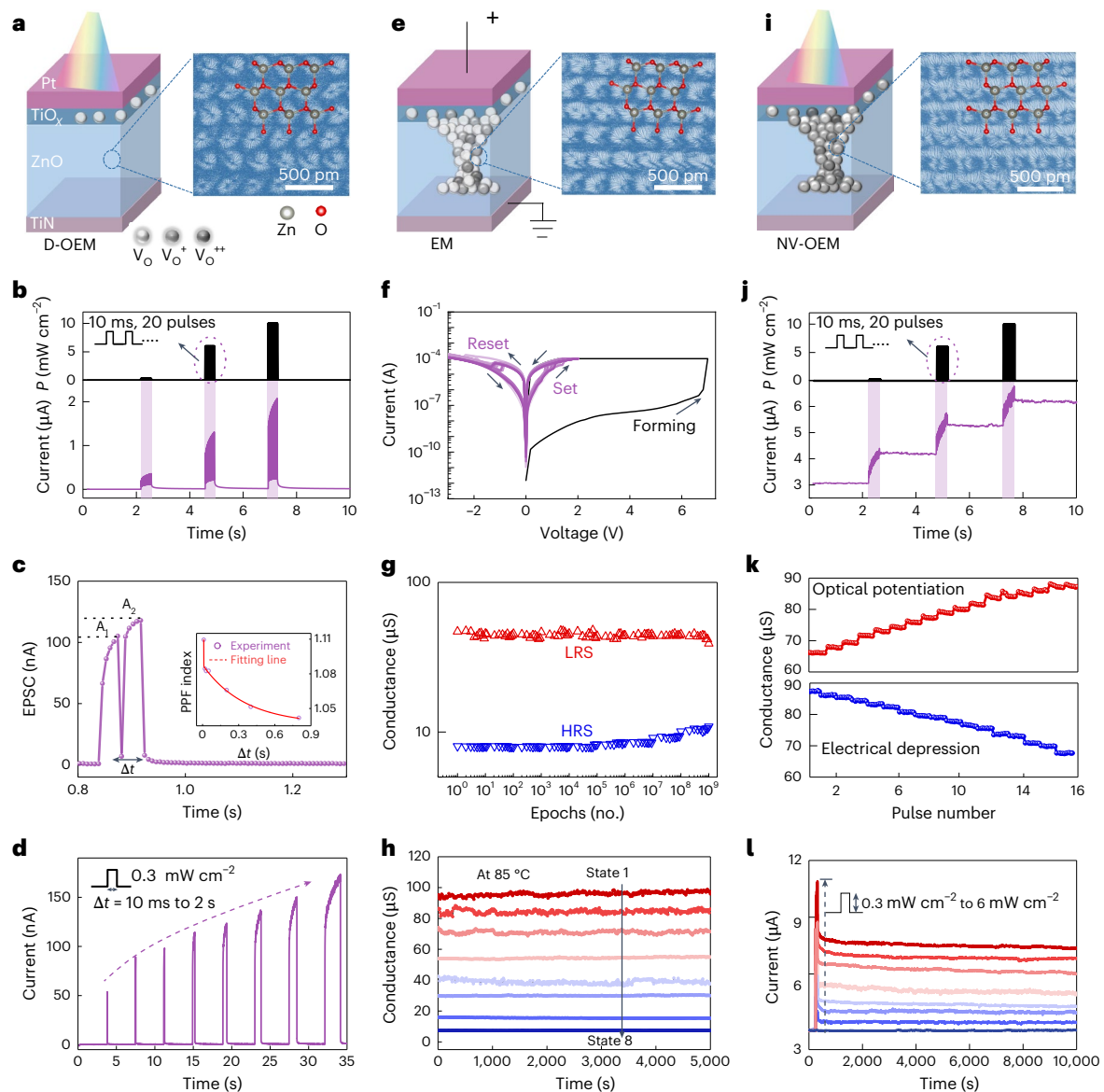


Fig. 2 | Configurable multi-functional OEM characteristics. **a**, A schematic illustration of the pristine D-OEM mode and the electric field intensity distribution within the ZnO from DPC-STEM. **b**, The EPSC D-OEM mode under three 20-pulse trains with different intensities (0.3 mW cm^{-2} , 6 mW cm^{-2} and 10 mW cm^{-2}). The light pulse width is 10 ms, and the read voltage is 0.3 V. **c**, EPSC responses by PPF (405 nm, 0.3 mW cm^{-2} and 50 ms). The EPSC of the second light stimulus (A_2) is higher than that of the first light stimulus (A_1). Inset: the PPF index (A_2/A_1) and the interval times (Δt) fitting result with a double-exponential function. **d**, The EPSC induced by different light pulse widths from 10 ms to 2 s under the same light intensity of 0.3 mW cm^{-2} . **e**, A schematic illustration of the forming process in the EM mode and an electric field intensity distribution within the conductive filament from DPC-STEM. **f**, Direct-current I - V characteristics of the EM mode. In the electroforming process, the voltage sweeps from 0 V to 7 V under a compliance current of 100 μA . When the voltages sweep from

0 to 2 V or -3 V, the conductance is set/reset to the LRS/HRS, respectively. **g**, Endurance measurements during 10^9 cycles under a pulse width of 100 ns with 3.5 V for set and -3.5 V for reset. **h**, A retention test over a long period of 5,000 s at 85°C for eight conductance levels. **i**, A schematic illustration of the NV-OEM mode after the optical process and the electric field intensity distribution within the conductive filament from DPC-STEM. **j**, The EPSC of the NV-OEM under three 20-pulse trains with different light intensities (0.3 mW cm^{-2} , 6 mW cm^{-2} and 10 mW cm^{-2}). The light pulse width is 10 ms. **k**, A schematic illustration of multi-state potentiation by optical stimulus (from $12 \mu\text{W cm}^{-2}$ to 9.6 mW cm^{-2} with 0.6 mW cm^{-2} steps, a wavelength of 405 nm and a pulse width of 10 ms) and depression by electrical stimulus (-1.2 V to -2.7 V, with 0.1 V steps and pulse width of 10 ms) for 16 pulses. **l**, The long-time retention characteristics of the NV-OEM under varying light intensities of 0.3 – 6 mW cm^{-2} with a pulse width of 2.5 s and pulse number of 15.

deviations in the electron beam, attributable to the oxygen vacancies (V_o and V_o^+) in bulk being ionized into V_o^+ and V_o^{++} owing to light irradiation. As shown in Fig. 2j, the non-volatile photocurrent can be adjusted by the consecutive light pulses with three different intensities. These non-volatile behaviours are analogous to long-term plasticity in biological synapses. We also study the response time of NV-OEM down to a minimum pulse width of 10 μs with an intensity of 0.03 mW cm^{-2} , as depicted in Supplementary Fig. 13. The underlying mechanism of the

NV-OEM is proposed as follows: when exposed to optical illumination, the TiO_x/ZnO interface defects are excited and oxygen vacancies (V_o and V_o^+) in bulk are ionized into V_o^+ and V_o^{++} (refs. 36,37), generating free electrons and increasing the device conductance (Supplementary Fig. 4k–m). When the optical illumination is removed, a portion of the electrons rapidly recombine with defects located at the TiO_x/ZnO interface and with oxygen vacancies at shallow defect energy levels. However, other electrons are prevented from recombining with

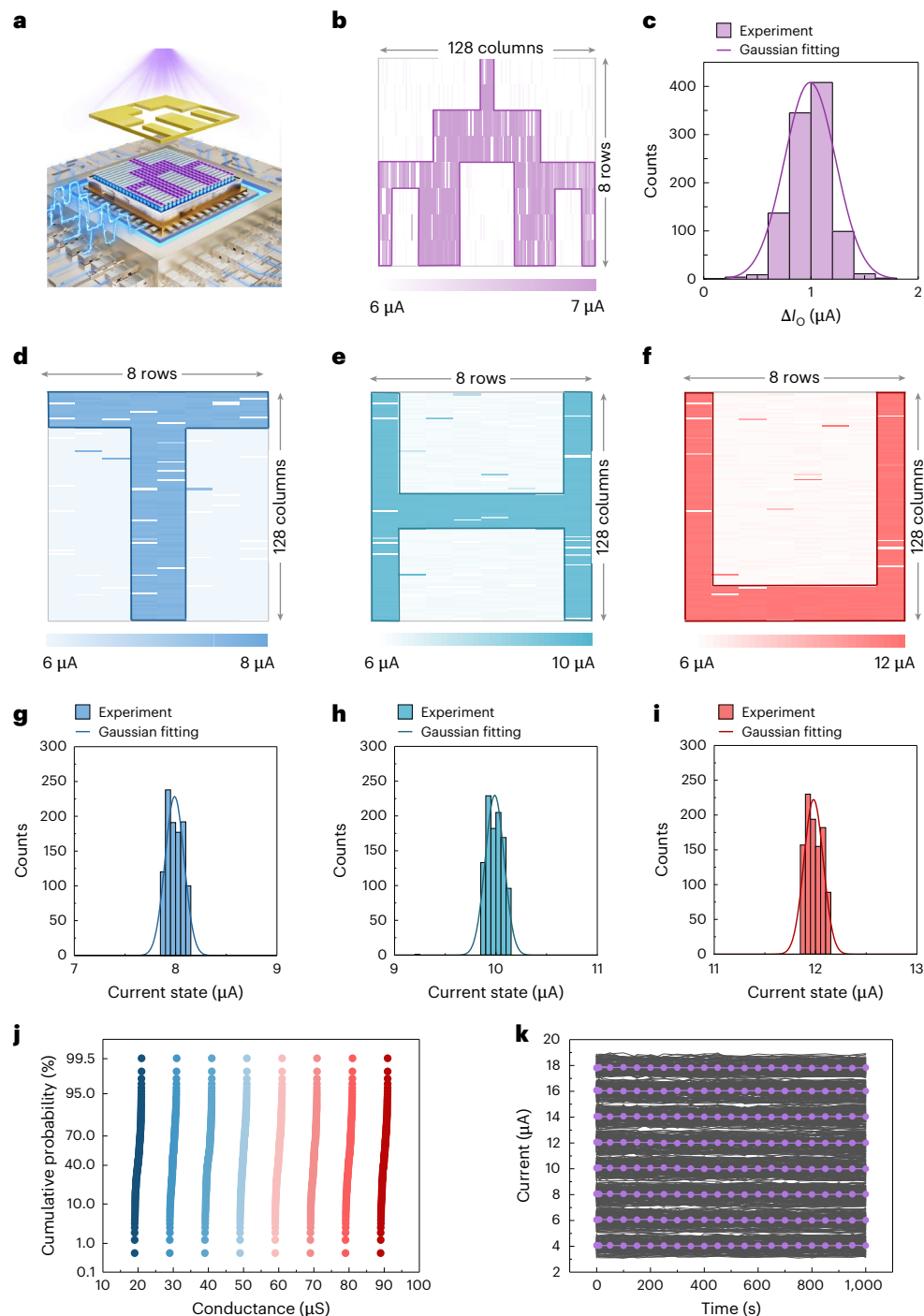


Fig. 3 | Image storage and pattern mapping in the OEM array. **a**, An illustration of the optical and electrical stimulus on the OEM array. **b**, The non-volatile storage of an image pattern (a gate icon) on the OEM array mapped by 405 nm light stimulus for 300 s with a pulse width of 200 ms and light intensity of 10 mW cm^{-2} . **c**, The corresponding distribution of the photocurrent response (ΔI_0). The line represents the result of Gaussian fitting. **d–f**, Mapping of the letters ‘THU’ with three different conductance targets of 8 μA (**d**), 10 μA (**e**) and 12 μA (**f**) under electrical stimulus (with a pulse width of 1 ms) using incremental step pulse programming with a read voltage of 0.2 V. **g–i**, The distributions of three different conductance states of 8 μA (**g**), 10 μA (**h**) and 12 μA (**i**) under electrical stimulus on the OEM array. Each of these current distributions exhibits a high degree of concentration. **j**, The cumulative probability distribution of the array with eight equally distributed conductance states. **k**, The read current distribution of eight levels as a function of time.

the deeply localized oxygen vacancies and are instead drawn by the drain terminal. This results in the majority of oxygen vacancy defect states in the bulk remaining unoccupied by electrons, as shown in Supplementary Fig. 4n. These V_{O}^{++} and V_{O}^+ provide an electron hopping pathway, thereby non-volatily increasing the conductivity after the light stimulus.

Furthermore, the non-volatile multi-level conductance could be reversibly adjusted by light stimulus (potentiation) and electrical stimulus (depression) by an incremental light intensity stimulus, as shown in Fig. 2k. This can achieve a relatively linear response, as shown in Supplementary Fig. 14, when compared with a stimulus of identical light intensity. Supplementary Fig. 15 demonstrates the results of

12 consecutive long-term potentiation and long-term depression cycles. The non-volatile characteristic of NV-OEM is evident by the long retention of $>10^4$ s for eight different states following the optical stimulus, as shown in Fig. 2l.

Leveraging oxygen vacancy conductive filaments to adjust defect energy level distributions enables precise control over diverse photoelectric response modes, which diverges from the photoelectric procedures reported in previous studies. Additionally, comparative analysis demonstrates that the OEM system operates with the advantages of low energy consumption and a wide spectral response (Supplementary Figs. 16 and 17).

The 128×8 1T1OEM array consists of 1,024 cells integrated on the backend of the line of Si CMOS circuits. The 128 rows can be accessed individually through a decoder (Supplementary Fig. 18 and Supplementary Note 1). Furthermore, the functionality of a 1T1OEM cell is further elucidated in Supplementary Fig. 19. These results demonstrate the array's capability for multi-bit programming and matrix–vector multiplication (MVM) operations. To verify the uniformity of the 128×8 1T1OEM array under the optical and electrical stimuli, we analyse the responses of EM and OEM in the array, as depicted in Fig. 3a. Figure 3b displays the non-volatile visual memory function in the OEM array. To accomplish this, we use a photomask with a transparent gate pattern during light stimulation. When the entire OEM array is exposed to the same laser stimulation, the photocurrent (ΔI_0) variation follows a Gaussian distribution (Fig. 3c), indicating the excellent uniformity of the optical stimulus.

Additionally, we create a conductance map of the letters ‘THU’ by assigning them to three distinct target conductance levels. A binary verification operation was utilized until the conductance reached the desired target levels ($I(T) = 8 \mu\text{A}$, $I(H) = 10 \mu\text{A}$ and $I(U) = 12 \mu\text{A}$). Detailed insights into the programming method can be found in Supplementary Note 2 and Supplementary Fig. 20. Figure 3d–i displays the results of our distinct target conductance mapping. Figure 3j illustrates the cumulative probability distribution of the 1 kb 1T1OEM array, which is programmed to eight uniformly distributed conductance states. Figure 3k shows the retention performance. These results suggest that the OEM array exhibits high uniformity and stability in response to electrical and optical stimulations, paving the road for the following demonstrations of in-sensor computing.

Image pre-processing by the NV-OEM-based sensing system

Image pre-processing plays a crucial role in visual information processing, which can suppress noise, enhance features and facilitate the subsequent processing tasks^{6,38}. Figure 4a displays the noise reduction and feature enhancement simulation using the NV-OEM-based sensing system. The handwritten digit images (Fig. 4b) are obtained by adding red–green–blue (RGB) Gaussian noise from 0 to 255 to the original modified National Institute of Standards and Technology (MNIST) dataset image. Supplementary Note 3 describes the construction of a convolution kernel for image pre-processing. The pre-processing of convolution kernels can significantly suppress background noise and highlight the image's features, as shown in Fig. 4c. In this demonstration, we employ the device model, proceeding with the training of the model weights in an offline manner. This simulation emulates the device's response to natural light filtering scenarios as shown in Supplementary Fig. 21, where the illumination signal is directly inputted. Following the pre-processing stage, the recognition of MNIST handwritten digit images is performed using a typical CNN with a multi-layer perceptron of $784 \times 300 \times 10$, as shown in Fig. 4d.

When an NV-OEM-based sensing system performs pre-processing, the recognition accuracy remarkably reaches $96.1 \pm 0.2\%$ after ten training epochs using the noisy dataset. The result is close to the recognition accuracy of the standard MNIST dataset without background noise,

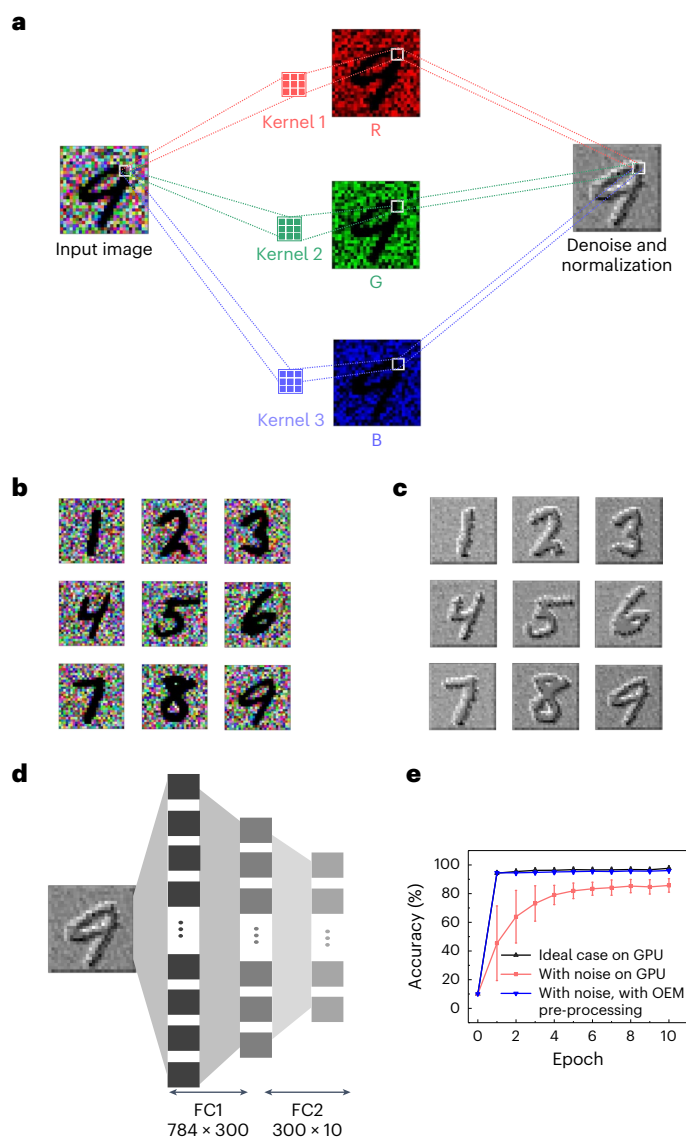


Fig. 4 | Image pre-processing using the NV-OEM-based sensing system. **a**, A schematic illustration of noise reduction and feature enhancement using the NV-OEM-based sensing system (construction of a convolution kernel size of 3×3 for the three RGB channels). **b**, Typical images with RGB Gaussian noise injected. **c**, The images after NV-OEM-based pre-processing. **d**, The recognition of MNIST handwritten digit images through a $784 \times 300 \times 10$ multi-layer perceptron, which is composed of two distinct fully connected layers (FC). **e**, Comparisons of the image recognition accuracy without noise, with noise but without pre-processing, and with noise and pre-processing by the NV-OEM-based sensing system. All values represent the mean \pm s.d. ($n = 5$).

as shown in Fig. 4e. In comparison, the recognition accuracy of the noisy dataset without pre-processing is only $85.7 \pm 4.7\%$. This approach ensures the accuracy and reliability of the NV-OEM-based sensing system with RGB noise in the background scene. We also perform the simulation of the recognition accuracy for training in Supplementary Note 4 and Supplementary Fig. 22, where both training and testing results achieved over 90% accuracy.

Object tracking by the hybrid OEM sensing system

Object tracking is a high-level sensory processing task in the human brain that involves different neuronal systems working together to locate a target accurately^{17,39,40}. Figure 5a displays image frames

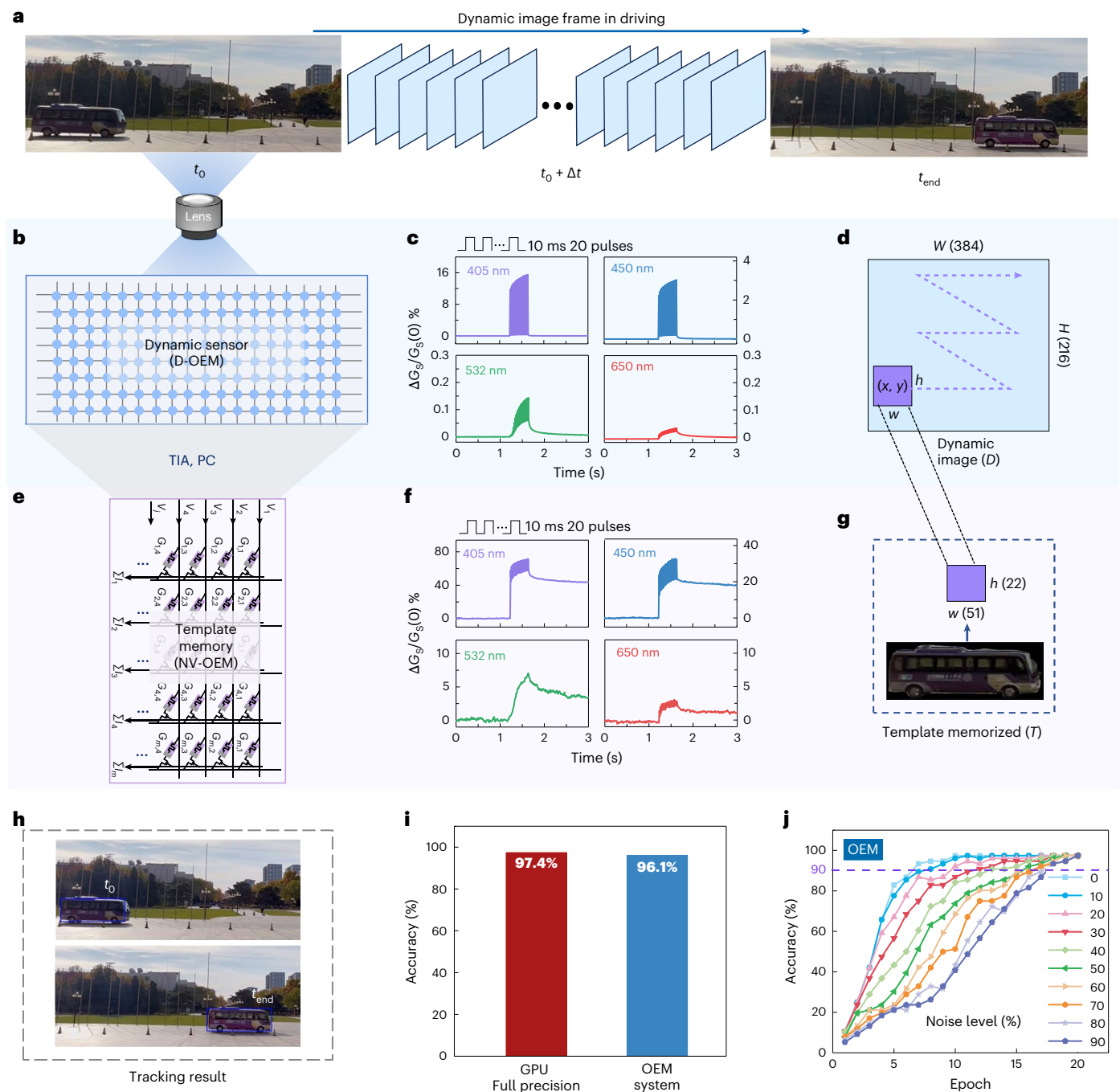


Fig. 5 | Object tracking by the hybrid D-OEM and NV-OEM sensory system.

a, An example of the dynamic image frame of a bus driving. **b**, A schematic of the D-OEM array that senses the dynamic image frame through optical signals. **c**, The conductance state (G_s) programming by a series of 20 light pulses with wavelengths of 405 nm, 450 nm, 532 nm and 605 nm at the same light intensity of 10 mW cm^{-2} in the D-OEM mode. **d**, The dynamic image (D) sensing by the D-OEM array with $W \times H(384 \times 216)$ pixels. **e**, A schematic diagram of the NV-OEM matrix circuit for the stored template, where the corresponding electrical signal is input after undergoing conversion via the D-OEM array, facilitated by transimpedance amplifier (TIA) and personal computer (PC) processing. V_1, \dots, V_j represents the

input voltage, and G represents the device conductance. **f**, The G_s programming using a series of 20 light pulses with wavelengths of 405 nm, 450 nm, 532 nm and 605 nm at the same light intensity of 10 mW cm^{-2} in the NV-OEM mode. **g**, A diagram of the template (T) storage with $w \times h(53 \times 22)$ NV-OEM pixels and the matching process using the TemplateMatchModes. **h**, Dynamic tracking results of the bus at the initial (t_0) and final (t_{end}) moments. **i**, The accuracy of the standard tracking algorithm by GPU and the hybrid OEM sensing system. **j**, The accuracy of the hybrid OEM sensing system with the noise level varied from 0% to 90%, where increasing numbers of training epochs are needed to achieve the same accuracy.

captured at different times, depicting the bus driving route from t_0 to t_{end} . The dynamic image sensor module comprises a D-OEM array (Fig. 5b) with a dynamical photo-response capability ranging from UV to visible light and high sensitivity to short wavelengths, especially for 450 nm and 405 nm light that are verified in Fig. 5c. Using the D-OEM

sensor array, we can perceive the external image in real time (Fig. 5d). The NV-OEM array (Fig. 5e) with non-volatile photo-response capability performs the template storage function as illustrated in Fig. 5f. As a result, the template of the bus is stored on the NV-OEM array by the photoconductivity as shown in Fig. 5g.

The tracking algorithm is described in Methods. The template-matching algorithm depends on the maximum value matched in the corresponding region, even if the response magnitudes between D-OEM and NV-OEM under the same light intensity are different, which does not affect the results of subsequent template-matching operations. On the basis of the high sensitivity of the D-OEM device to short wavelengths, it can be regarded as an efficient filter capable of extracting the bus region from the RGB values. As a result, by combining OEM system with tracking algorithms, the precise location of the bus can be tracked. In the end, Fig. 5h shows the results of bus tracking after the OEM-based template-matching algorithm, and the complete tracking video is shown in Supplementary Video 1.

To address the limitations imposed by manufacturing equipment, we use the OEM characteristics measured in the experiment to construct the large-scale software model for tracking tasks. To meet the demands of real-world video frame rates when transmitting dynamic photoelectric signals, our experiment demonstrates a video tracking screen that operates at a frame rate of 25 fps. Supplementary Note 5 and Supplementary Table 1 indicate that the fastest processing frequency of our system is almost 0.1 MHz.

The performance of the OEM sensor architecture and the standard tracking algorithm run by a graphics processing unit (GPU) are compared. As shown in Fig. 5i, the OEM sensing system achieves a high accuracy of 96.1%, which is almost identical to the GPU result of 97.4%. Additionally, we introduce different levels of Gaussian noise (ranging from 10% to 90%) to the OEM and GPU system during the bus driving process, which both increase the number of training epochs required to achieve the same level of accuracy, as shown in Fig. 5j (OEM results) and Supplementary Fig. 23 (GPU results). Meanwhile, our OEM architecture can also accommodate multi-target tracking and re-recognition functions in the target tracking task, as shown in Supplementary Video 2.

Fully OEM-based RC system

In the demonstration of OEM-based RC, multiple OEMs are used to construct a physical RC system for classifying human motion data in the NTU-RGB dataset (Fig. 6a). We construct a fully OEM-based RC framework with a D-OEM layer of virtual dynamic sensor reservoir nodes and a read-out layer based on EM in one hardware, as shown in Fig. 6b. The human motion signals input is described in Methods and Supplementary Figs. 24 and 25. Owing to the transformation of spatiotemporal information into light-encoded input into the sensor reservoir, which does not require training, the RC shows the advantage of reduction in training costs and network size. A total of 18 D-OEM devices are used to receive the 18 channel signals in parallel. An example (B_5 , M_4) of the input signal and D-OEM responses is shown in Fig. 6c. The state vectors can be obtained by measuring the output photocurrent of the D-OEM. Given the same input sequence, multiple parallel devices can produce different state vectors due to the device-to-device variation, which has been proven beneficial to the RC system performance as it increases the state richness²³. More experiment details

about the 18 channel optical signals and the response are provided in Supplementary Fig. 26.

In the training step for the read-out layer, the state vectors are collected from the output current and the labels are used in the ridge regression algorithm to calculate the output weights according to the training protocol of RC. Next, the output weights are mapped to the output layer in RC, which consists of a crossbar array of EM modes. It can naturally perform MVM in an array, making it suitable for the read-out layer. The training process is described in the Supplementary Note 6. Owing to the limited scale of the array, we utilize five identical 1k arrays to map the 720×5 weights by using 90×8 devices on each array. After training, the colour graphs with 720×5 weights obtained through software simulation and OEM hardware experiment turned out to be highly consistent within the permissible range of errors. Figure 6d shows the corresponding target weights, hardware weights and errors as a result of conductance mapping. The read conductance indicates that the weight matrix is successfully mapped onto the OEM in a non-volatile manner with acceptable error for implementing the output layer of the OEM-based RC. The mapping result is then used to calculate the accuracy for the human motion recognition task, such as drink water, throw, sit down, kick or jump up.

Then, in the inference process, an array composed of 18 D-OEM devices is used for the reservoir layer where the high-dimensional vector signals are generated by light stimulation. The signal collection and voltage conversion are carried out through an off-chip testing system. Subsequently, these vectors are fed into the 1k 1T1OEM array for MVM assisted by the on-chip peripheral circuits and a memory testing system. This inference process is performed in parallel, with vectors being input through the eight bit lines and yielding output currents that represent the MVM results. The details of training experimental procedures employed in human motion recognition demonstration are elaborated in Supplementary Fig. 27 and Supplementary Note 6. Figure 6e compares the obtained normalized MVM results and the theoretical values, indicating that the OEM array can faithfully implement MVM computations. The confusion matrix for the GPU system classification results shows an accuracy of 93.2% (Supplementary Fig. 28). Figure 6f illustrates the results for the OEM hardware, which achieves a recognition accuracy of 91.2%. Figure 6g presents a comparison of the recognition accuracy between the full-precision GPU simulation and the fully OEM-based in-sensor RC hardware.

Conclusions

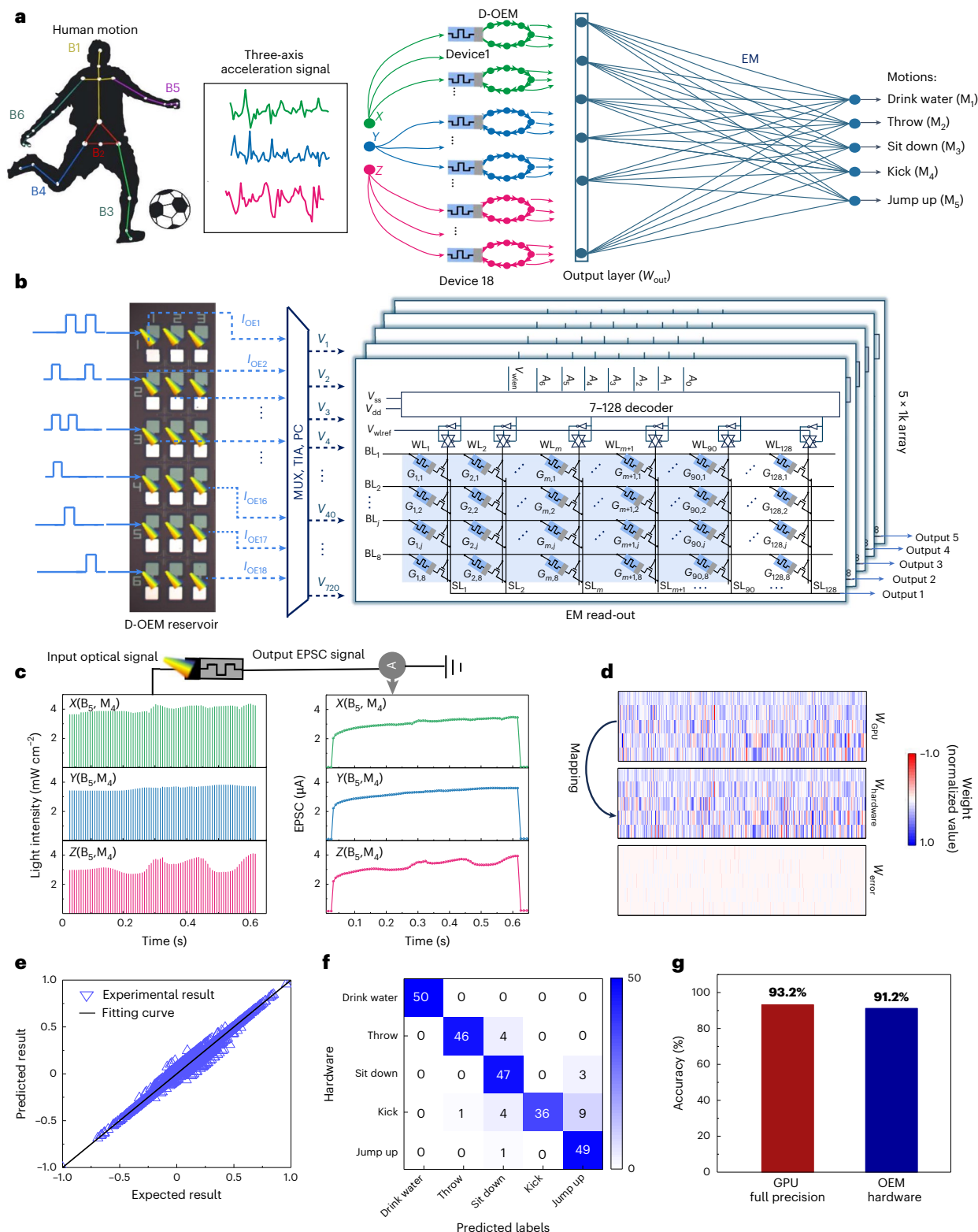
We have developed a multi-mode OEM based on Pd/TiO_x/ZnO/TiN that embodies EM, D-OEM and NV-OEM functionalities in a single device. The D-OEM can be transformed into the NV-OEM through an electric-forming stimulus, creating conductive filaments and introducing deep-level oxygen vacancy defects to enhance the optoelectrical memory function. Using this device, a highly integrated 128×8 1T1OEM array is fabricated and exhibits high uniformity and stability in response to electrical and optical stimulations. Furthermore, we have

Fig. 6 | Fully OEM-based in-sensor RC system. **a**, Fully OEM-based RC for human motion recognition. Six skeleton groups on the human body are labelled (B_1 – B_6). In parallel, 18 D-OEM devices in the reservoir layer receive 18 light channel signals (six skeleton points, each with x, y and z coordinates, that is, $6 \times 3 = 18$). These signals are processed using multiplexer (MUX), TIA and PC, generating 720 high-dimensional vector signals (owing to each of the 18 channels collecting 40 virtual nodes, leading to a state vector length of $18 \times 40 = 720$). These vectors are then fed into a 1T1OEM array for MVM to recognize five types of human motion: drink water (M_1), throw (M_2), sit down (M_3), kick (M_4) and jump up (M_5). W_{out} , output layer weight matrix. **b**, The circuit diagram of the fully OEM chip. This chip comprises an 18 D-OEM array constituting the reservoir computing layer, which outputs the current (I_{OE}) and converts it into the input voltage (V). Additionally, five identical 1k EM arrays implement the readout layer (mapping the 720×5 weights using 90×8 devices on each array). Each array is equipped with decoding

circuit (A_0 – A_6) to select rows, bit lines (BL_1 – BL_8), source lines (SL_1 – SL_{128}), and word lines (WL_1 – WL_{128}), source line voltage control and output multiplexing, enable signal (V_{wen}); transistor gate voltage control (V_{wref}), power (V_{dd}) and ground (V_{ss}) signals. **c**, Examples of the input signals for kick (M_4) and the skeleton groups of B_5 in which the coordinate information is encoded. Each optical pulse has a width of 10 ms. The D-OEM output EPSC is monitored using a constant read voltage of $V_{read} = 0.2$ V. **d**, Distributions (720×5) of the target conductance weights (W_{GPU}), the OEM hardware weight ($W_{hardware}$) and the error values (W_{error}). **e**, A comparison of the normalized MVM results from the array and the theoretically expected values. **f**, The confusion matrix of the recognition accuracy, based on the fully OEM architecture. **g**, A comparison of the recognition accuracy after training for the GPU full-precision simulation and the fully OEM-based in-sensor RC hardware.

demonstrated the advantages of multiple OEM by implementing three representative tasks, from simulation to experiment in a fully hardware OEM system. For low-level sensory processing, the sensing system based on NV-OEM mode can effectively simulate image pre-processing functions, such as noise reduction and feature enhancement, improving the recognition accuracy from 85.7% to 96.1%. For high-level processing, the hybrid system, employing both D-OEM and NV-OEM, effectively mimics object tracking, achieving an impressive accuracy of 96.1%. Additionally,

we showcase a fully OEM-based in-sensor RC system for human motion recognition, where the D-OEMs serve as the physical reservoir layer and the EM implements the output layer. This RC system enabled human motion classification to reach a GPU-equivalent recognition accuracy of 91.2%. Compared with existing research (Supplementary Table 2), the key advance of our OEM system is the large-scale integration of 1T1OEM with configurable modes in one hardware. Furthermore, we have evaluated the computational energy consumption of the OEM



system in comparison with GPU, as depicted in Supplementary Note 7 and Supplementary Table 3. The results also suggest that reducing the light pulse width can further decrease the energy consumption of the OEM system by more than 20 fold. Thus, further optimization of the OEM can be carried out in the future by (1) optimizing the top electrode using transparent materials or plasmons to increase the light absorption rate, (2) utilizing heterogeneous structures or P–N junctions to enhance the built-in electric field and improve the optical responsivity and (3) optimizing the configuration of our testing equipment. The OEM system developed in this work provides a cost-effective and energy-efficient platform for implementing high-accuracy in-sensor neuromorphic computing at the edge, enabling next-generation artificial vision chips.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41565-024-01794-z>.

References

- Zhou, F. et al. Optoelectronic resistive random access memory for neuromorphic vision sensors. *Nat. Nanotechnol.* **14**, 776–782 (2019).
- Wu, N. Neuromorphic vision chips. *Sci. China Inf. Sci.* **61**, 060421 (2018).
- Zhou, F. & Chai, Y. Near-sensor and in-sensor computing. *Nat. Electron.* **3**, 664–671 (2020).
- Dai, S. et al. Emerging iontronic neural devices for neuromorphic sensory computing. *Adv. Mater.* **35**, 2300329 (2023).
- Du, J. et al. A robust neuromorphic vision sensor with optical control of ferroelectric switching. *Nano Energy* **89**, 106439 (2021).
- Li, G. et al. Photo-induced non-volatile VO₂ phase transition for neuromorphic ultraviolet sensors. *Nat. Commun.* **13**, 1729 (2022).
- Jang, H. et al. An atomically thin optoelectronic machine vision processor. *Adv. Mater.* **32**, 2002431 (2020).
- 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).
- Zhao, R. et al. A framework for the general design and computation of hybrid neural networks. *Nat. Commun.* **13**, 3427 (2022).
- Wang, Y. et al. Optoelectronic synaptic devices for neuromorphic computing. *Adv. Intell. Syst.* **3**, 2000099 (2021).
- Zhang, J., Dai, S., Zhao, Y., Zhang, J. & Huang, J. Recent progress in photonic synapses for neuromorphic systems. *Adv. Intell. Syst.* **2**, 1900136 (2020).
- Song, S. et al. Recent progress of optoelectronic and all-optical neuromorphic devices: a comprehensive review of device structures, materials, and applications. *Adv. Intell. Syst.* **3**, 2000119 (2021).
- John, R. A. et al. Reconfigurable halide perovskite nanocrystal memristors for neuromorphic computing. *Nat. Commun.* **13**, 2074 (2022).
- Wang, T. et al. Reconfigurable neuromorphic memristor network for ultralow-power smart textile electronics. *Nat. Commun.* **13**, 7432 (2022).
- Pi, L. et al. Broadband convolutional processing using band-alignment-tunable heterostructures. *Nat. Electron.* **5**, 248–254 (2022).
- Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
- Zhang, Z. et al. All-in-one two-dimensional retinomorphic hardware device for motion detection and recognition. *Nat. Nanotechnol.* **17**, 27–32 (2022).
- Sun, L. et al. In-sensor reservoir computing for language learning via two-dimensional memristors. *Sci. Adv.* **7**, eabg1455 (2021).
- Wu, X. et al. Wearable in-sensor reservoir computing using optoelectronic polymers with through-space charge-transport characteristics for multi-task learning. *Nat. Commun.* **14**, 468 (2023).
- Lao, J. et al. Ultralow-power machine vision with self-powered sensor reservoir. *Adv. Sci.* **9**, 2106092 (2022).
- Zhong, Y. et al. A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing. *Nat. Electron.* **5**, 672–681 (2022).
- Liang, X. et al. Physical reservoir computing with emerging electronics. *Nat. Electron.* **7**, 193–206 (2024).
- Moon, J. et al. Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat. Electron.* **2**, 480–487 (2019).
- Portner, K. et al. Analog nanoscale electro-optical synapses for neuromorphic computing applications. *ACS. Nano.* **15**, 14776–14785 (2021).
- Hu, L. et al. All-optically controlled memristor for optoelectronic neuromorphic computing. *Adv. Funct. Mater.* **31**, 2005582 (2021).
- Tan, H. et al. An optoelectronic resistive switching memory with integrated demodulating and arithmetic functions. *Adv. Mater.* **27**, 2797–2803 (2015).
- Chen, J. Y. et al. Dynamic evolution of conducting nanofilament in resistive switching memories. *Nano Lett.* **13**, 3671–3677 (2013).
- Simanjuntak, F. M., Panda, D., Wei, K. H. & Tseng, T. Y. Status and prospects of ZnO-based resistive switching memory devices. *Nanoscale Res. Lett.* **11**, 368 (2016).
- Xu, N. et al. Characteristics and mechanism of conduction/set process in TiN/ZnO/Pt resistance switching random-access memories. *Appl. Phys. Lett.* **92**, 232112 (2008).
- Zhou, Z., Pei, Y., Zhao, J., Fu, G. & Yan, X. Visible light responsive optoelectronic memristor device based on CeO_x/ZnO structure for artificial vision system. *Appl. Phys. Lett.* **118**, 191103 (2021).
- Wang, T.-Y. et al. Reconfigurable optoelectronic memristor for in-sensor computing applications. *Nano Energy* **89**, 106291 (2021).
- Wang, W. et al. CMOS backend-of-line compatible memory array and logic circuitries enabled by high performance atomic layer deposited ZnO thin-film transistor. *Nat. Commun.* **14**, 6079 (2023).
- Wang, Z. et al. Vacancy driven surface disorder catalyzes anisotropic evaporation of ZnO (0001) polar surface. *Nat. Commun.* **13**, 5616 (2022).
- Lanza, M. et al. Recommended methods to study resistive switching devices. *Adv. Electron. Mater.* **5**, 1800143 (2019).
- Kuzum, D., Yu, S. & Wong, H. S. Synaptic electronics: materials, devices and applications. *Nanotechnology* **24**, 382001 (2013).
- Russo, P., Xiao, M., Liang, R. & Zhou, N. Y. UV-induced multilevel current amplification memory effect in zinc oxide rods resistive switching devices. *Adv. Funct. Mater.* **28**, 1706230 (2018).
- Oh, I., Pyo, J. & Kim, S. Resistive switching and synaptic characteristics in ZnO/TaON-based RRAM for neuromorphic system. *Nanomaterial* **12**, 2185 (2022).
- Seo, S. et al. Artificial optic-neural synapse for colored and color-mixed pattern recognition. *Nat. Commun.* **9**, 5106 (2018).

39. Yilmaz, A., Javed, O. & Shah, M. Object tracking. *ACM Comput. Surv.* **38**, 1–45 (2006).
40. Wang, S. et al. Networking retinomorph sensor with memristive crossbar for brain-inspired visual perception. *Natl Sci. Rev.* **8**, nwaa172 (2020).

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Heyi Huang^{1,2,3}, Xiangpeng Liang¹, Yuyan Wang^{1,2}✉, Jianshi Tang^{1,2}✉, Yuankun Li¹, Yiwei Du¹, Wen Sun¹, Jianing Zhang⁴, Peng Yao¹, Xing Mou¹, Feng Xu¹, Jinzhi Zhang⁴, Yuyao Lu¹, Zhengwu Liu¹, Jianlin Wang⁵, Zhixing Jiang¹, Ruofei Hu¹, Ze Wang¹, Qingtian Zhang¹, Bin Gao^{1,2}, Xuedong Bai⁵, Lu Fang^{2,4}, Qionghai Dai², Huaxiang Yin³, He Qian¹ & Huaqiang Wu^{1,2}✉

¹School of Integrated Circuits, Beijing Innovation Center for Integrated Circuits, Tsinghua University, Beijing, China. ²Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China. ³Integrated Circuit Advanced Process R&D Center and Key Laboratory of Fabrication Technologies for Integrated Circuits, Institute of Microelectronics of Chinese Academy of Sciences, Beijing, China. ⁴Department of Electronic Engineering, Tsinghua University, Beijing, China. ⁵Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, China. ✉e-mail: wangyuyan@tsinghua.edu.cn; jtang@tsinghua.edu.cn; wuhq@tsinghua.edu.cn

Methods

The 1T1OEM array fabrication

The OEM array was fabricated through a back-end process after utilizing a standard 130 nm Si CMOS process in a commercial foundry to fabricate the transistors and peripheral circuits. After receiving the wafer from the foundry, it was cleaned with 50:1 hydrofluoric acid to remove surface oxides. Then, a 30-nm-thick TiN bottom electrode was deposited using magnetron sputtering on top of the drain terminal of the Si transistors. After that, a 50-nm-thick ZnO layer was deposited using magnetron sputtering, followed by the deposition of Ti (5 nm) and Pd (5 nm) by electron-beam evaporation. Finally, the OEM stack was etched to complete the 1T1OEM array. The fabrication process flow of the 1T1OEM cell is shown in Supplementary Fig. 29.

Electrical and optical measurement

To evaluate the cell and array, electrical characterizations were carried out using a Keithley 4200 semiconductor parameter analyser and a Cascade Summit 12000 probe station, while optical characterizations were measured using 405 nm, 450 nm, 532 nm and 650 nm laser excitations. For the single cell, we measured the output current using a read voltage of 0.3 V. The 1k bit 1T1OEM array was assessed by the SpeedCury ST2516 memory tester system and probe cards, mapping the array with a read voltage of 0.2 V to monitor the EPSC. Supplementary Fig. 30 illustrates the experimental set-up for achieving in-sensor computing using the OEM chip.

Structural analysis

The cross-sectional sample was prepared by focused ion beam technique (FIB Helios G5 DualBeam, Thermo Fisher), and the TEM characterization was performed by TEM (JEOL, ARM300F) operated at 200 kV. X-ray photoelectron spectroscopy was carried out on a Shimadzu AXIS SUPRA+ under a monochromatic AlK α (photon energy 1,486.6 eV). The absorption spectra of films on quartz were measured using a Cary 5000 UV–vis near infra-red, Agilent and Excalibur 3100 Varian spectrophotometer, with wavelengths ranging from 200 nm to 800 nm.

Tracking algorithm

First, the NV-OEM mode stores the bus's initial template matrix T ($w \times h$), where w and h represent the width and the height in pixel dimensions, respectively. The bus template is mapped onto the NV-OEM array through the optical signals. Since the photocurrent generated is non-volatile in the exposure array, the change of the photoconductivity on the corresponding array is also memorized. Second, the D-OEM mode is used to sense the dynamic image matrix D ($W \times H$) during the driving. Transimpedance amplifier converts the dynamic current generated in the volatile mode into voltage. This voltage is then divided into column vectors of equivalent size to the template. Subsequently, these vectors are input into the non-volatile array for matrix multiplication. Finally, the similarity between each region and the template is determined by TemplateMatchModes according to the computer, where the region with the maximum value is defined as the best matching location (R)

$$R(x, y) = \text{Max} \left(\sum_{x', y'} (T(x', y') \times D(x + x', y + y')) \right), \quad (2)$$

where $x' = 0 \dots W - 1$, $y' = 0 \dots H - 1$ are the summation ranges.

Human motion signal input

The original data consist of three-axis coordinates of skeletons for various human motions. In this work, we select five types of motion from the dataset, including drink water (M_1), throw (M_2), sit down (M_3),

kick (M_4) and jump up (M_5) and six skeleton groups of the human body including B_1 , B_2 , B_3 , B_4 , B_5 and B_6 , with each motion recorded at a frame rate of 100 fps. In the pre-processing step, as shown in Supplementary Fig. 24, the coordinate data are encoded into a light sequence with a wavelength of 405 nm and pulse width of 10 ms, which is injected into the D-OEM to induce a photo-response for high-dimensional mapping. In the D-OEM mode, the minimum light intensity for a detectable resolution can be as low as $8 \mu\text{W cm}^{-2}$, as shown in Supplementary Fig. 25. To achieve reliable signal collection with peripheral circuits, the actual demonstration adopt the light intensities within the realistic visible light range of $1\text{--}10 \text{ mW cm}^{-2}$.

Data availability

The data that support the findings of this study are available from the corresponding authors on request. More data are also presented in the Supplementary Information. Source data are provided with this paper.

Code availability

All the codes that support the findings of this study are available from the corresponding authors upon reasonable request.

Acknowledgements

This work was in part supported by National Natural Science Foundation of China 92264201 (J.T.), 62025111 (H.W.) and 62174095 (Y.W.), National Key R&D Programme of China 2021ZD0109901 (L.F.), China Postdoctoral Science Foundation 2021M701845 (H.H.), the XPLOER Prize (H.W.), Tsinghua University Initiative Scientific Research Programme and the Centre of Nanofabrication, Tsinghua University. We are also grateful to J. Chen from the University of Zurich and J. Tang from Boston College for their valuable suggestions on the manuscript.

Author contributions

H.H., Y.W. and J.T. conceived and designed the experiments. H.H. contributed to the OEM system fabrication. W.S., H.H., J.W. and X.B. contributed to the TEM analysis. H.H., X.M., Z.J. and R.H. participated in the measurements. Y. Li performed a simulation of image noise reduction under the supervision of Q.Z.; Jianing Zhang and Jinzhi Zhang performed a simulation of object tracking under the supervision of L.F. and Q.D.; and X.L., Y. Li and H.H. performed experiments on RC. F.X. and Y. Lu provided theoretical support. Y.D., P.Y., Z.L., Z.W., B.G., H.Y. and H.Q. analysed the data and discussed the results. H.H., X.L., Y.W. and J.T. wrote the manuscript. All authors discussed the results and commented on the manuscript. Y.W., J.T. and H.W. supervised the project.

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/s41565-024-01794-z>.

Correspondence and requests for materials should be addressed to Yuyan Wang, Jianshi Tang or Huaqiang Wu.

Peer review information *Nature Nanotechnology* thanks Jang-Sik Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.