Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Memristive blinking neuron enabling dense and scalable photonically-linked neural network

Bojun Cheng

bocheng@hkust-gz.edu.cn

The Hong Kong University of Science and Technology (Guangzhou) https://orcid.org/0000-0001-5551-4242

Yue ZHOU

The Hong Kong University of Science and Technology (Guangzhou)

Yuetong Fang

The Hong Kong University of Science and Technology (Guangzhou)

Raphael Gisler

ETH Zurich

Hongwei Ren

The Hong Kong University of Science and Technology (Guangzhou)

Haotian Fu

The Hong Kong University of Science and Technology (Guangzhou)

Zelin Ma

The Hong Kong University of Science and Technology (Guangzhou)

Yulong Huang

The Hong Kong University of Science and Technology (Guangzhou)

Renjing Xu

The Hong Kong University of Science and Technology (Guangzhou) https://orcid.org/0000-0002-0792-8974

Alexandre Bouhelier

Laboratoire Interdisciplinaire Carnot de Bourgogne, CNRS UMR 6303 Université de Bourgogne https://orcid.org/0000-0002-0391-2836

Juerg Leuthold

ETH Zurich https://orcid.org/0000-0003-0111-8169

Article

Keywords:

Posted Date: November 25th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5489562/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Electronics on January 8th, 2026.
See the published version at https://doi.org/10.1038/s41928-025-01529-5.

Memristive blinking neuron enabling dense and scalable photonically-linked neural network

Abstract

The rapid growth of artificial intelligence (AI) has driven a surging demand for powerful architectures whose capabilities scale with network size and dense neuron connections. However, the large footprint of CMOS neurons and the constraints on electric routing are significantly hindering the scale-up of the number of artificial neurons and their synaptic connectivity. Inspired by the biological neurons' intricate and dense network structure, we introduce a 3-dimensional (3D) connectivity. Here, we demonstrate a 3D neural network (NN) with atomic scale memristive blinking neuron (MBN) linked by photons free from electrical routing constraints. Our MBN features a minimal footprint of 170×240 nm^2 and emits photon pulses upon integrating a critical number of incoming electrical spikes, eliminating the need for the bulky peripheral circuit readout and electrical wiring for transmitting the signals. Such MBN not only offers the functionality of a spiking neuron but also constitutes a CMOS-compatible planar light source for intra-neuron communication, overcoming the bottleneck of plenary CMOS technology. The functionality of the MBN is demonstrated in two different photonically-linked NN applications. First, we test the MBN on a four-class classification task within the Google Speech dataset and show 91.51% accuracy. We then show its operation in a high-density artificial neuron array with a pitch of 1 µm, pushing the boundaries of optical diffraction, realizing an unprecedented neuron density on par with the human brain. An MNIST classification task is performed with a 92.27% accuracy. Our new photonically-linked NN significantly increases the neuron density as well as scalability, showing great potential for realizing next-generation AI hardware.

Introduction

Deep learning enables significant advancements in a wide range of applications[1], from natural language processing and image analysis to autonomous driving. This accompanies a surging demand for extensive hardware complexities[2]: the computation of large AI models has doubled every 3-4 months, far outpacing the scaling of integrated circuits[3] since 2012. As network size continues to grow, CMOS transistors encounter critical challenges in scaling down, nearing their physical limits[4-7]. Moreover, challenges with crossing electrical path[8] and increased heat generation[9] further constrain the densely packed CMOS circuits. The aspiring goal would be to achieve a system on par with the human brain, containing tens of billions of neurons and connections within a dense structure to perform complex cognition and learning functions[10] while consuming only tens of watts. Inspired by the human brain, there is a trend for developing neuromorphic chips for the next-generation AI accelerators, taking advantage of innovative devices[11] and paradigms.

Memristors are gaining popularity as next-generation devices due to their versatile functionalities and compact size[12-14]. Redox-based memristors, in particular, are of significant interest because of their filamentary switching behavior, which is achieved with the movement of only a few atoms within a compact device[15,16]. While non-volatile memristors are commonly employed as artificial synapses to mimic synaptic weights[17-19], volatile memristors can function as artificial neurons, effectively replicating the leak, integrate-and-fire behavior[20]—replacing dozens of CMOS transistors in the process[21,22]. When an applied voltage exceeds a certain threshold, volatile devices can rapidly switch from a high to a low resistance state and relax back to the low resistance state once the voltage is removed[23,24]. Probing a memristor's resistance

state requires dedicated peripheral circuits, such as trans-impedance amplifiers (TIA) or charge-based capacitors[25], which significantly increase the footprint of sub-μm² memristors to hundreds of μm² [26,27]. These systems also face trade-offs between energy consumption, speed, and accuracy due to parasitic capacitance and excess electrical noise[28]. The large-scale neuron-to-neuron connection is further restricted by the intricate electrical routing protocols and the high manufacturing cost in the 3D stacked chip interconnect process.

To overcome the performance trade-offs and minimize the peripheral circuit footprint, utilizing optical signals to transmit and propagate neuron states presents a promising alternative[29]. Unlike electrical signals, photons are not subject to parasitic capacitance-induced performance trade-offs. Photons can propagate in direct line-of-sight[30] or through various photonic links[31], reducing crosstalk and enabling more flexible neuron arrangements, particularly in larger arrays[32,33]. Photonic links also have the potential to simplify the manufacturing process and lower the costs associated with 3D stacked chip interconnects. For photonic links, both photodetectors and photon sources are essential. The large-scale integration of photodetectors with sub-micron pitch is already well established in compact cameras used in mobile electronics[34,35], but there are not yet suitable on-chip photon sources for neural links. Scalable on-chip light sources such as LED and OLED still feature sizes of tens of μm^2 and require high current levels for light emission, whereas memristors operate at sub-microampere levels[36,37]. More advanced light sources, such as quantum dot technologies[38,39] require delicate and stable voltage conditions that are against the drastic changing operating states of neurons. Recently, an atomic-scale memristive light source (APS) has been reported[40], capable of emitting photon bursts during the resistive switching process. This type of light-emitting device gives the potential for integrating on-chip light sources with artificial neurons, bridging the gap between photonic links and memristive neuromorphic systems.

Here, we further exploit the memristor's transient emission characteristic to introduce an atomic-scale memristive blinking neuron (MBN). This compact device uniquely facilitates direct optical output from neurons through two key features: First, serving as a compact neuron while simultaneously generating neuron output optically. Second, it achieves the integrate-and-fire function within one single device along with emitting a burst of photons, which upon activation the device generates as an optical output. The dual function is unified within operating conditions eliminating the bulky electrical peripheral circuits and electrical wiring for amplifying, reading, and transmitting a neuron's output. As such, MBNs integrate the photonic link with the inherently compact nature of memristors, shrinking the neuron and photonic interconnect pitch to 1 μm. Second, the MBN operates based on a novel driving paradigm–photon emission from electrical current temporal variation (dI/dt). The high on/off ratio of memristors allows for substantial signal margins even at moderate operating currents, which is advantageous for implementing neural networks with a low power budget. Leveraging these two advantages, we demonstrate the applicability of the MBN to build a photonically-linked SNN for practical tasks. The MBN-based architecture achieves a 91.51% classification accuracy on Google speech and a 92.27% accuracy on MNIST, respectively.

Device demonstration

Optical and electrical characteristics of MBN. The MBN reported here is realized with a Ag/PMMA/Ag metal-insulator-metal (MIM) memristive switching in-plane junction. The fabrication details are given in Supplementary I. An MBN's conceptual illustration is depicted in Fig. 1a and a scanning electron microscope (SEM) image of a device is shown in Fig. 1b. The MBN uses two symmetrical triangular plasmonic nanoantennas separated by a gap to enhance the light emission efficiency. The bow-tie-shaped antenna triangle features a compact width of 170 nm and a length of 240 nm. It also serves as the switching electrode. The plasmonic antenna design, simulation, and measurement details are given in Supplementary II. Photons are emitted during the memristive switching process, with emission intensity (I_{ph}) positively correlated to the derivative of electrical current dI/dt. The light-emitting filamentary volatile switching behavior is characterized by current-voltage (*I-V)* sweeping experiment, shown in Fig. 1c. Once the applied

voltage exceeds the device's threshold, an abrupt increase of current is observed, accompanied by photon emission (see Fig. 1c inset ii), and the MBN goes into the low resistance state (LRS). Upon reaching the compliance current, the current remains stable, and the light emission ceases (see Fig. 1c inset iii). The 5 μA compliance current prevents the device from electrical damage. Once the voltage is reduced below the holding threshold, the current drops and the device returns to the high resistance state (HRS). We verify the reliability of device volatile switching through 1000 consecutive cycles, see Supplementary VI. We tentatively attribute the light emission to electroluminescent defects introduced during the formation of the conductive filament[40,41]. The luminescence mechanism and switching principle are discussed in more detail in Supplementary IV. A microscopic image of a single device in operation is zoomed in in Fig. 1d. Photons are emitted from the switching region and are confined to a diffraction-limited area with a diameter of about 0.75 μm (FWHM). The temporal evolution of the light intensity is detailed in a close-up view recorded by the avalanche photodiode (APD), illustrated in Fig. 1e: a burst of photons is generated upon the abrupt change of current, reaching a peak photon count of 233 binned over 10 μs, corresponding to an optical power of 5.43 pW. A broad emission spectral ranging from 780 to 920 nm (FWHM) is observed during the electroluminescence test, shown in Fig. 1f.

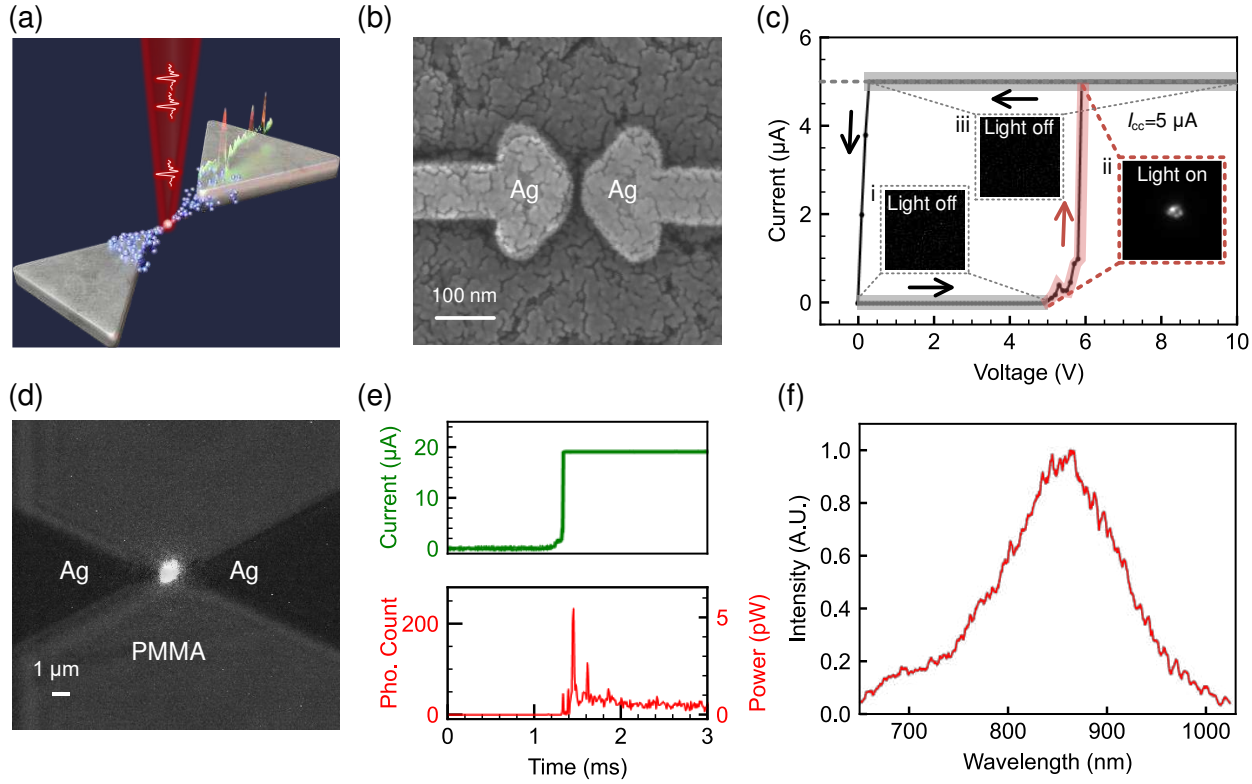


Fig. 1 | Optical and Electrical Characteristics of Memristive Blinking Neuron (MBN). a, Conceptual illustration of the memristive blinking neuron (MBN) with plasmonic nanoantenna. The red light cone illustrates the emission of photon spikes correlated to current change. **b**, SEM image of the MBN`s Ag antenna structure. A dielectric protective layer is deposited to cover the chip and to prevent surface charging, which results in the coarse grain boundary in the SEM image. **c**, The *I-V* sweep curve with a compliance current of 5 μA. Inset: electroluminescence images captured during the three stages of the *I-V* sweep. **d**, The electroluminescence image captured by the CMOS camera overlaid with an optical transmission image of the device, showing the location of the emission between the two electrodes. **e**, MBN's current response and time-resolved photon emission upon applying an 8 V, 100 ms bias. Upper: the output current (green curve). Bottom: the photon counts binned every 10 μs (red curve). **f**, Electroluminescence spectrum measured with a spectrometer.

Transient-mode electroluminescence. The transient-mode photon emission is further investigated by examining the relationship between current and optical activity. As seen in Fig. 2a, under a 10 V bias for 50 ms, the MBN exhibits a fluctuating current (green) accompanied by bursts of photons (red). We

introduce the current change rate dI/dt to correlate the photon emission with the current variation. The rate is displayed in Fig. 2a bottom (black), with a clearer view provided in the zoom-in plot in Fig. 2b. A strong correlation between the photon emission and dI/dt is evident: higher photon counts correlate to a larger value of $|dI/dt|$. Once the current stabilizes at the compliance level, the photon emission diminishes, marking the end of transient-mode electroluminescence.

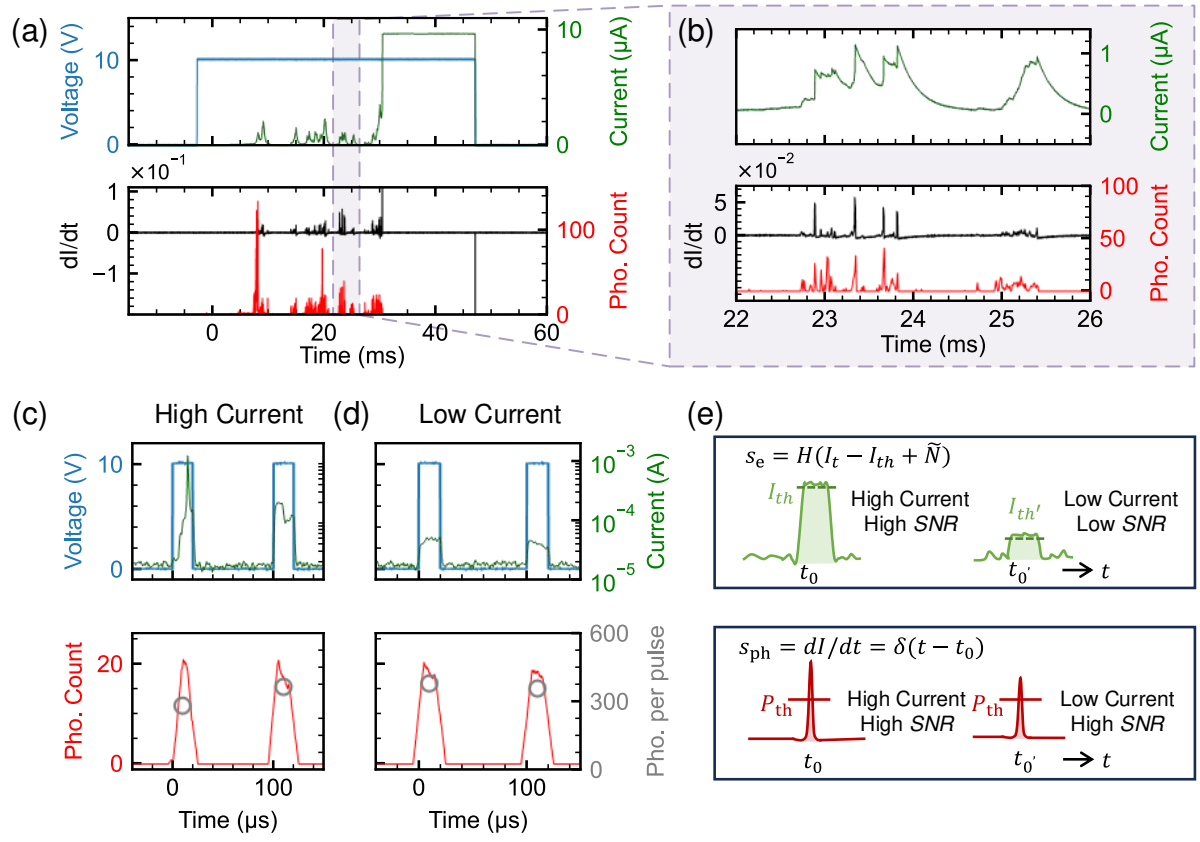


Fig. 2 | Transient-mode electroluminescence of Memristive Blinking Neuron (MBN). a, The transient electroluminescent behavior under a 10 V, 50 ms bias. Upper: the applied voltage bias (blue) with the output current (green). Bottom: the photon counts binned every 100 μs with the current change rate dI/dt. **b**, Zoomed-in plot of the data from panel a where the correlation of current change and photon emission is detailed. **c.** Operation of the MBN under high current conditions with a 20 μs short pulse at 10 V. Upper: The applied voltage bias (blue) alongside the output current (green). Lower: Photon counts binned every 10 μs (red) and per pulse (grey). **d,** Operation of the MBN under low current conditions with the same 20 μs pulse at 10 V. Upper: The applied voltage bias (blue) with the output current (green). Lower: Photon counts binned every 10 μs (red) and per pulse (grey). **e.** An illustration of electrical and photonic spikes under high current and low current scenarios. Electrical spikes occur above a specific current threshold, with higher currents yielding greater margins at the cost of higher power consumption. Photonic spikes, in comparison to electrical spikes, triggered by dI/dt, can maintain prominent signal intensity in both high- and low-current levels, supporting efficient detection and minimizing power-related trade-offs. P_{th} is the photon count threshold we set to binarize the photonic spike, here we adopt 5 photon count per pulse in our experiment.

The MBN's photonic readout is more agile and energy-efficient compared to conventional electrical readout. In conventional artificial neuron, switching is intrinsically determined by a threshold current I_{th}[42]. We summarize the spike behavior by a equation $s_e = H(I_t - I_{th} + \tilde{N})$, where $H(\cdot)$ denotes a Heaviside step function and $\tilde{N}$ represents the noise of current. This approach presents a trade-off in signal margin, readout speed, and energy consumption: higher current facilitates faster readouts and improved signal margins, but leads to increased power consumption. In contrast, the MBN's photonic spike intensity is related to dI/dt. The total photon emission during the switching is thus calculated as $\int dI/dt|dt = I_{on}/I_{off}$. Thanks to the high current on-off ratio (I_{on}/I_{off}) of the memristor, even a 90% reduction of I_{on} still yields ample current on/off ratio and high photon count, allowing operation in low power mode. This is experimentally illustrated

in Fig. 2c and 2d with high and low current scenarios with a 20 µs pulse at 10 V. Remarkably, the devices at low current levels emit photon counts comparable to those achieved at higher currents, despite the latter exceeding a tenfold difference. A more detailed discussion of the relationship between emission intensity, current levels, and the current change ratio can be found in Supplementary III. We further model this in Fig. 2e: when the current exceeds the threshold (at $t = t_0$), an electrical spike signal $s_\mathrm{e} = 1$ is generated. The photonic spike is modeled as $s_\mathrm{ph} = dI/dt = \delta(t - t_0)$. With discretized timestep, this can be simplified to $s_\mathrm{ph} = 1$ when $t = t_0$. This alignment of photonic and electrical spikes elegantly simplifies the model while serving as a foundational element for constructing neural networks with MBN.

Investigation of leaky integrate-and-fire process. The MBN not only aligns its optical and electrical output but also closely mimics the leaky integrate-and-fire (LIF) behavior, which is essential for neuromorphic computing. The optical readout enables the possibility to transmit signals to the next neuron via photons. Fig. 3a conceptually illustrates how our MBN integrates stimuli from pre-synaptic input to generate a transient photon pulse that mimics action potentials. The photon emission serves as the output signal and relay information to the next neuron once a sufficient threshold is reached[43]. As presented in Fig. 3b, the MBN integrates the stimuli consisting of a series of 8 V, 500 Hz pulses. When a critical number of pulses, N_crit, is accumulated, the MBN transits to LRS, producing the electrical spikes s_e (green line). Simultaneously, bursts of photonic spikes s_ph are measured by an APD (red lines). As such, we name the memristive neuron "blinking". From here, we eliminate peripheral circuits to probe the memristor's state and carry out all subsequent experiments with only an optical readout of the neuron by an external photodetector (either an avalanche photodetector (APD) or CMOS image sensor) at ambient temperature, as detailed in Supplementary VII.

To assess and extract the MBN's leak integrate-and-fire characteristic, we launch two pulse test groups with various amplitudes V_pulse and guard interval t_guard, respectively. We first stress the MBN with various pulse amplitudes. From Fig. 3c, it can be seen that a higher voltage leads to a faster build-up of the virtual membrane potential $U_\mathrm{mem}[t]$, and the MBN can be activated with fewer pulses, mirroring the natural neurons' integration process. In the next experiment, we examine the leaky characteristics with 5 V, 500 µs pulses followed by guard times of different durations, as depicted in Fig. 3d. For longer guard intervals, more pulses are required to activate the MBN, reflecting the gradual "leak" of accumulated charge, similar to biological neuron[44]. The MBN's behavior can be discretized by a classic LIF model detailed in Eq. (1), with all parameters derived from experiments in Fig. 3c and 3d:

$$U_\mathrm{mem}[t] = \left(1 - \frac{\Delta t}{\tau}\right) U_\mathrm{mem}[t - \Delta t] + \Delta U[t], \tag{1}$$

where Δt denote the timestep with $\Delta t = t_\mathrm{pulse} + t_\mathrm{guard}$. τ denotes the membrane potential leak constant, which was determined to be 29.7 ms in our MBN model. Here, the $\Delta U[t]$ gives the increase of the membrane potential due to the stimulus of synapses. It is nonlinearly related to the voltage by a mapping weight function $w(V)$. $\Delta U[t]$ writes,

$$\Delta U[t] = w(V) \cdot s_\mathrm{in}[t] = \left(\alpha \cdot e^{\beta \cdot V} + \gamma\right) \cdot s_\mathrm{in}[t], \tag{2}$$

where $s_\mathrm{in}[t]$ represents the incoming spikes received at time t. In our experiment, we fit the measured weight $w(V)$ by the exponential relation in Eq. (2) with values $\alpha = 3.88 \times 10^{-8}$, $\beta = 1.58\ \mathrm{V}^{-1}$, and $\gamma = 0.067$.

After extracting the function in Eq. (2), the synaptic weight can be realized by a corresponding amplitude V for the incoming pulses with the mapping function $w(V)$. The detailed process of fitting of $w(V)$ and τ are in Supplementary VIII. When the virtual membrane potential exceeds a certain firing threshold $\mathrm{U_{th}} \equiv 1$, the neuron generates an output spike, denoted by $s_\mathrm{out}[t]$,

$$s_\mathrm{out}[t] = H(U_\mathrm{mem}[t] - U_\mathrm{th}), \text{with } H(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x > 0 \end{cases} \tag{3}$$

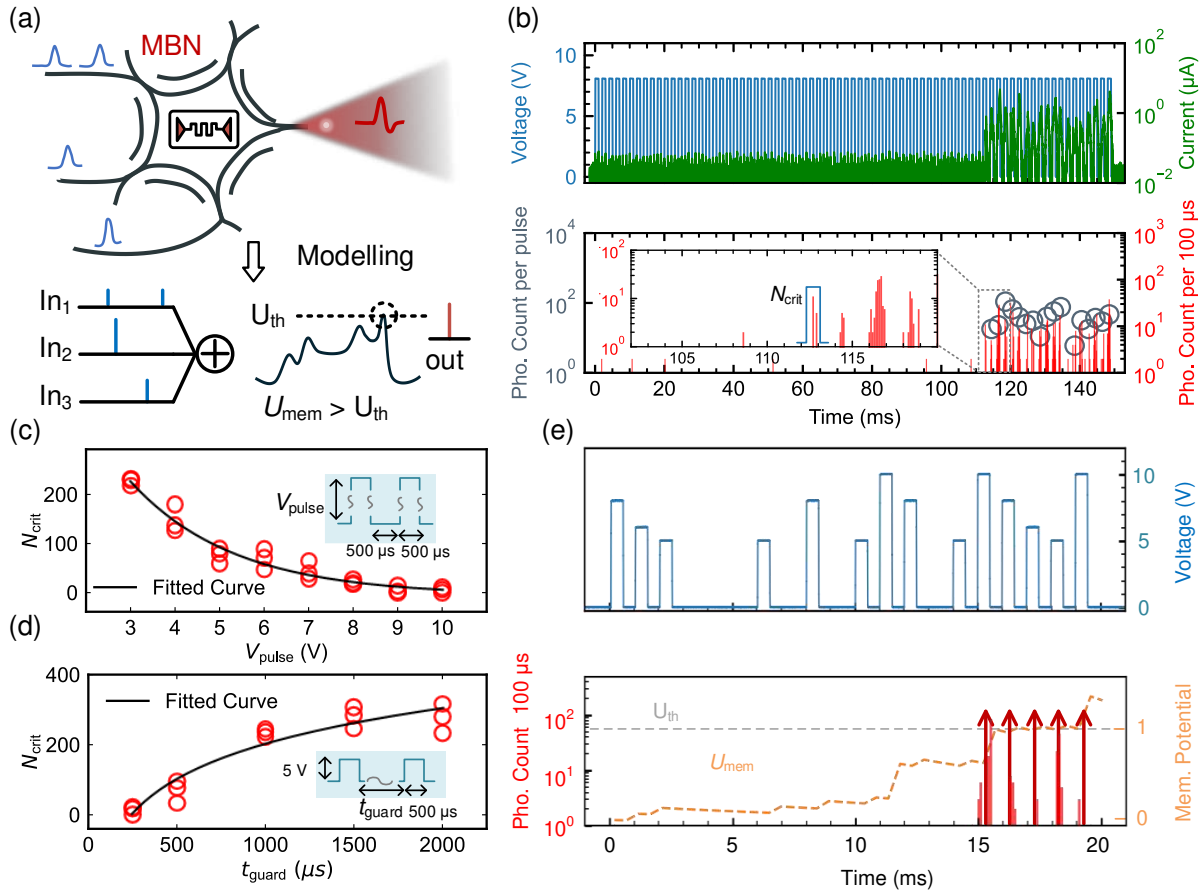


Fig. 3 | MBN's Leaky Integrate-and-Fire (LIF) Process. a, A schematic depiction of the MBN's integrating pre-synaptic input stimuli and generating a transient photon pulse and the abstract model of the MBN. This photon emission acts as the output signal, relaying information to the next neuron upon reaching the threshold. **b**, The leaky integrate-and-fire process of a single MBN device with a series of input voltages (500 Hz, 8 V). Upper: the input voltage and the output current. Bottom: corresponding photon emission with the photon counts binned per pulse (left) and per 100 μs (right). **c**, Leaky, integrate-and-fire characteristic of MBN tested with 1000 Hz, 500 μs pulses at various voltage amplitude. With higher amplitudes, fewer pulses are required for activation. This experiment is performed with input pulses at different amplitudes. **d**, Leaky, integrate-and-fire characteristic of MBN tested with 5 V, 500 μs pulses at various guard intervals. With longer guard intervals, more pulses are required for activation. **e**, MBN firing demonstration activated by a voltage train (top) accumulated from 4 synaptic input channels with different weighted inputs. The virtual membrane potential is calculated based on the extracted model and is plotted in the bottom, orange dashed line. The photon spike (measured in red) is observed whenever $U_{\mathrm{mem}} > \mathrm{U_{th}}$, validating the model.

Building on the extracted LIF model of the MBN from Fig. 3c and 3d, we experimentally verified the model by constructing a simple one-layer network consisting of one MBN connected to four synapses. The stimulus of input spike trains $s_{\mathrm{in}}[t]$ is combined across all input channels i for each timestep $\sum_i w_i \cdot s_{\mathrm{in},i}[t]$, the spike inputs and weight value of four channels are detailed in Supplementary V. This combined stimulus is converted into voltage trains of varying amplitudes according to $V[t] = w^{-1}\left(\sum_i w_i \cdot s_{\mathrm{in},i}[t]\right)$. The upper panel of Fig. 3e displays the voltage train fed into the MBN. We demonstrate the MBN's corresponding photon emission process in the bottom panel of Fig. 3e. Whenever the virtual membrane potential U_{mem} (orange dashed line) reaches the firing threshold $\mathrm{U_{th}}$ (gray dashed line), the MBN emits photons (red). Due to the stochastic nature of the resistive switching, the number of emitted photons varies with each cycle. To address this cycle-to-cycle variance, we binarize the MBN's photon output into a photon spike when the threshold of 5 photons per pulse is reached. This binarization, represented by arrows, ensures robustness against cycle-to-cycle variance and eliminates noise from the APD dark count, providing reliable photonic inter-layer interconnection.

Photonically-linked Spiking Neural Network with MBN

To demonstrate the applicability of the MBN in 3D optically connected neural networks, we constructed a simple and efficient photonically-linked spiking neural network (SNN) for a real-time Keyword Spotting (KWS) task with four different classes. We thereby leverage the energy-efficient benefit of SNNs originating from their sparse and event-driven nature[45].

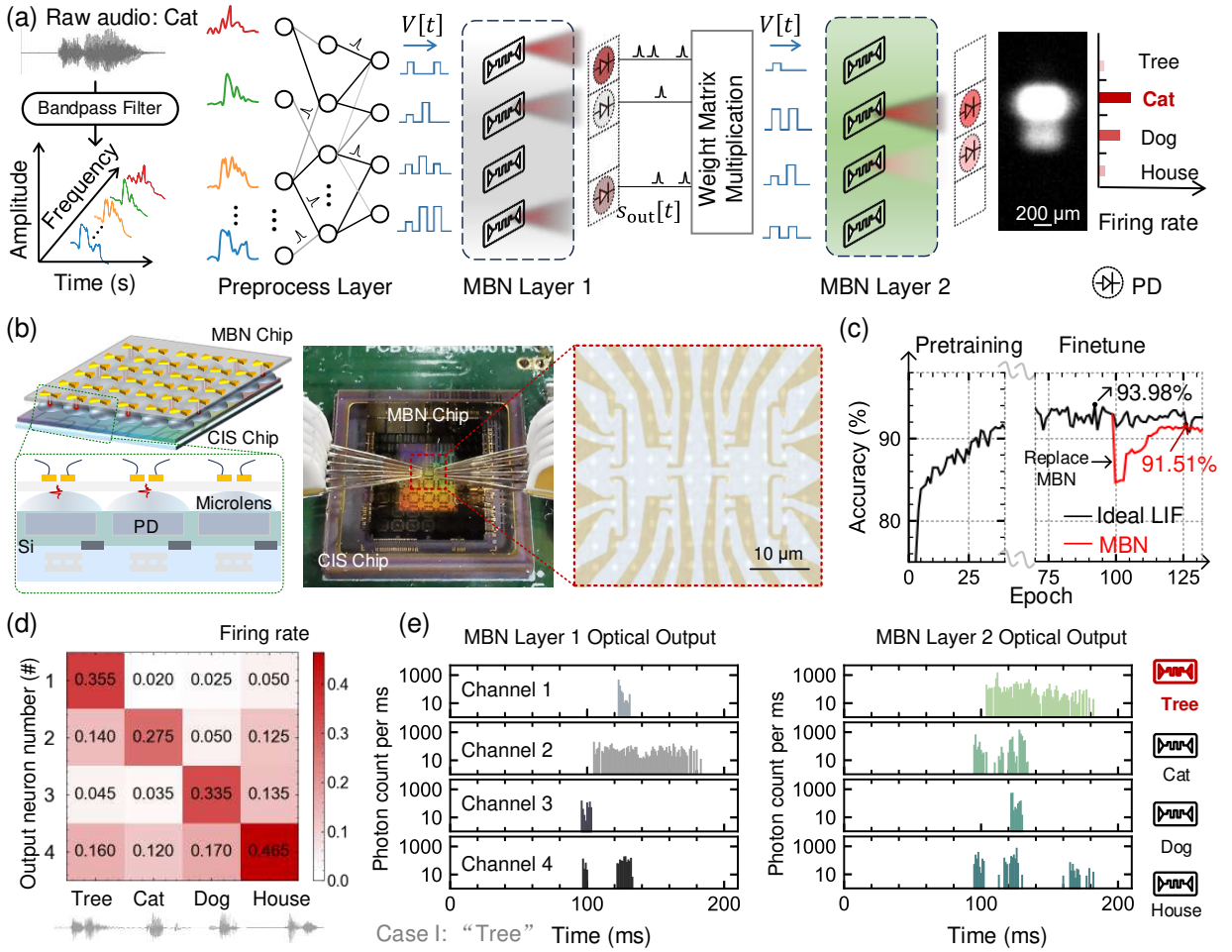


Fig. 4 | Photonically-linked Spiking Neural Network (SNN) for Keyword Spotting Task. a, The photonically-linked SNN architecture. The audio sequence is bandpass filtered and converted into spike trains by the preprocessing layer. The spike trains are sent into the MBN Layers 1 and 2, each containing four MBNs. The output of the MBN Layer 2 is directly probed by a CMOS image sensor (CIS) stacked below. **b,** Left: Conceptual illustration of a 3D photonic link achieved by integrating a MBN chip onto a CIS chip. Middle: Experimental setup with the MBN chip directly stacked on top of the CIS chip, without any electrical interconnections. Right: Overlay microscope image showing the CIS pixel aligned with the MBN neuron array. **c**, The network's training accuracy vs training epoch. We first pretrain with an ideal LIF neuron model (solid black line) and then replace the ideal LIF neuron with the experimental MBN model (solid red line) to finetune the learning process. The network finally achieves an accuracy of 91.51% with the MBN model. **d**, The inference results from the output layer with four different labels as conducted by MBN devices. The average firing rate is highest for the correct classes. **e**, An example of the inference process: "Tree". The figures illustrate the original photon counts from the first and second MBN layers. The highest photon firing rate is obtained for the "Tree" as predicted.

In practice, the photonically-linked SNN for KWS comprises a preprocessing convolutional module and 2 fully-connected layers with MBNs, as depicted in Fig. 4a. Taking the token "Cat" as an example, the audio input of the word "Cat" is first decomposed into 32-time series signals through band-pass filter banks. The bandpass signals are then converted into binary spikes through spiking neuron layers (more details are shown in Supplementary IX). The preprocessing convolutional module extracts the high dimensional

features from these spike sequences and merges them into 4-channel voltage trains with a timestep of 1 ms. After the preprocessing convolutional module, the spikes are subsequently fed into two fully connected layers, each built by four MBNs, see Fig. 4a right. MBNs take voltage train as input and generate photon spikes as output. Notably, the signal is transmitted photonically between the two neuronal layers, eliminating otherwise intricate electrical intra-chip connections. The second MBN layer serves as the output layer. Ideally, the output layer with the correct class is emitting most photon spikes. To demonstrate the applicability of MBNs for intra-chip optical interconnects, we stack directly the MBN chip on a CMOS image sensor (CIS) chip without any lenses or electrical connections between the two chips. The conceptual setup is shown in Fig. 4b (left), with a photograph of the actual setup in Fig. 4b (middle). A microscope view of a representative neuron array stacked on the CIS is presented in Fig. 4b (right). Further details of the setup can be found in the Supplementary VII. The images captured by the CIS clearly showcase the electroluminescence of the MBN within the array (Fig. 4a, right), demonstrating a direct intra-chip photonic link between neurons.

The network is trained offline. The training curves are shown in Fig. 4c. To achieve better performance, the network has first been pretrained with an ideal LIF neuron model (black), then finetuned with the MBN neuron model (red) that aligns with the measured quantities in Eq. (1). The training reaches a final accuracy of 91.51% (red) using the Google Speech dataset in a four-class classification (more details of the training process are given in Supplementary IX). The small difference between the red and black curves proves that the optical MBN can reach a performance that is comparable to the one obtained by an ideal LIF model. Experimental inference results of 4 labels are conducted, and the firing rate of all four MBNs in the second layer are recorded and summarized in Fig. 4d, all giving correct results. A detailed inference example in Fig. 4e shows that the correct label produces the most photon spikes, the complete inference examples are demonstrated in Supplementary IX.

Ultra-compact Artificial Neuron Array

To make next-generation neural networks scalable, artificial neuron arrays must feature an ultra-compact footprint. Thus, a small size is vital. This requirement can be seen for instance in spiking convolutional neural networks, where the number of neurons scales quasi-quadratically with the input image resolution n[46]. Here, we show that by taking advantage of an advanced nanofabrication technology removing electrical readout circuits, we can shrink the device pitch down to 1 μm for the MBN array. With such an ultra-compact footprint, our MBN array can arrange neurons in a 2D form and realize a highly dense 2D image information transmission through optical spikes. By emitting light signals directly from the chip surface, our MBN-based photonically-linked SNN can realize a more compact and efficient 3D information processing architecture.

To demonstrate the potential of MBN as building blocks of high-density inter-layer connects in neuromorphic hardware, we construct a convolutional spiking neural network comprising multiple layers of MBN arrays to apply the MNIST dataset for image recognition. At first, we simulate the whole neural network with all its layers by relying on the extracted MBN model. In the next step, we confine the offline processing to the first few layers and experimentally implement the last layer consisting of a 2×2 MBN array. As a showpiece, the last 2×2 MBN array with a 1 μm pitch has been realized on-chip to illustrate the direct transmission of neuron's output with optical spikes.

The MBN-based photonically-linked SNN consists of a convolution-based preprocessing layer followed by three spiking convolutional layers. As each single MBN is restricted to emitting unsigned photon spikes, we implemented a differential convolution strategy with two neurons acting as excitatory and inhibitory neurons, respectively, to represent either positive or negative output spikes, see Fig. 5a bottom (details are provided in Supplementary X). After passing both the excitatory and inhibitory MBN layer through a Heaviside function $H(\cdot)$, we attain the feature map of the layer, as illustrated in Fig. 5a top. Such an approach is essential to achieve inhibitory behavior and attain high accuracy in tasks with complex features. With the extracted MBN model, we train the photonically-linked SNN and achieved a 92.27% training

accuracy on the MNIST dataset for 10-category classification in the simulation. Then, we verify the network with the last layer comprising the fabricated MBN array. Two examples of input numbers “2” and “3” are illustrated in Fig. 5b and 5c. The excitatory and inhibitory MBN array in the third layer are recorded with a CMOS image sensor, illustrated in red and blue, respectively. The photon emission pattern matches the simulation result, highlighted in the excitatory (red) and inhibitory (blue) box.

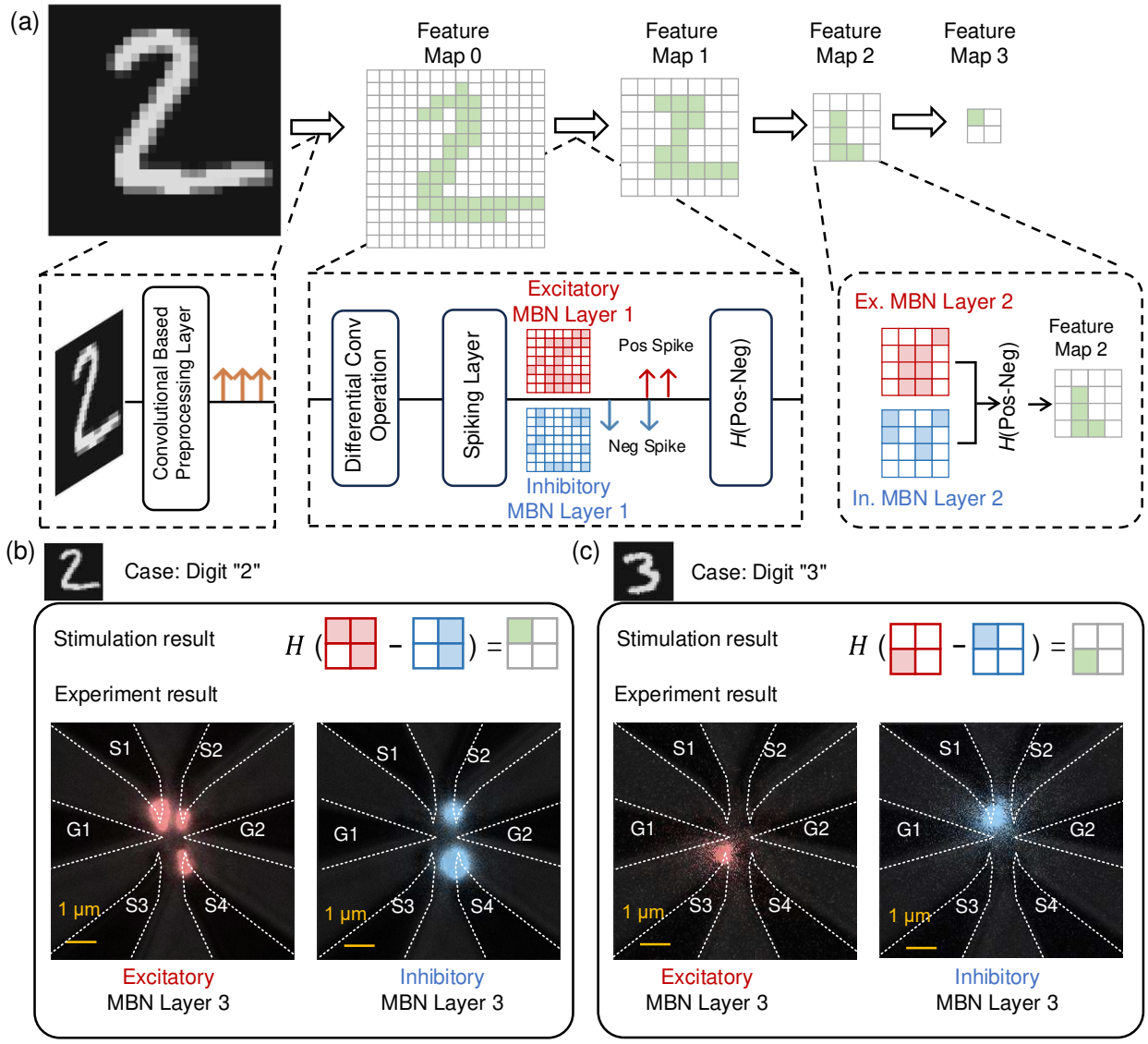


Fig. 5 | Photonically-linked Spiking Neural Network for Handwritten Digits Classification Task. a, The architecture of the photonic-linked spiking neural network with a convolution-based preprocessing layer followed by three spiking convolutional layers. The output of each layer is a feature map. The convolutional layer adopts a differential convolutional block with excitatory and inhibitory neurons. **b & c**, The MBN layer 3 is visualized with its 2 × 2 excitatory MBN array and a 2 × 2 inhibitory MBN array for input digits “2” and “3” respectively. The images are captured by a CMOS camera and consist of an overlay with the transmission microscopic image of the electrodes. The layout of the circuit is represented by dashed white lines. Labels: S_i stands for signal electrode i and G_n stands for ground electrode n.

To substantiate our claim, we analyze our neuron array density and compare it to existing artificial neuron implementations. By organizing MBNs into a peripheral circuit-free array, and considering an optically resolvable emission area for individual photon sources, we achieve a groundbreaking density of 1 μm^2 per nonlinear computing device. This density not only approaches the fundamental limit imposed by optical diffraction but also significantly surpasses the capabilities of both CMOS-based and other optical neural

network architectures, as summarized in Table 1. The effective density also surpasses the effective density of conventional electrical memristor-based neurons when considering the peripheral circuit area for the memristor's signal readout. With a demonstrated pitch of 1 μm², our design theoretically enables a staggering density of one million neurons per mm², approaching the neuron cellular density of the human brain[47].

Table 1 Comparison of the Occupied Area Across Different Non-linear Activation Function Units

Compared work	Realization method	Operation domain	Activation Function	Area
Analog Spiking Neuron[48]	Analog CMOS neuron	Electrical	LIF	34 μm^2
ROLLS neuromorphic processor[22]	Analog CMOS neuron	Electrical	LIF	917 μm^2
Hybrid CMOS-NEMS LIF neuron[49]	CMOS-NEMS neuron	Electrical	LIF	80 μm^2
All-optical spiking NN[50]	Ring resonator	Optical	LIF	~14400 μm^2
Multilayer nonlinear optical NN[51]	Image intensifier	Optical/Electrical/Optical	Sigmoid	> 36 μm^2
ACCEL photoelectronic chip[52]	Photodiode	Optical/Electrical	Power Function	1225 μm^2
On-chip Photonic deep NN[53]	Microring modulator	Optical	ReLU	225 μm^2
Our work	**Memristive blinking neuron**	**Electrical/ Optical**	**LIF**	**1 μm^2**

Conclusion

In this study, we introduce a novel memristive blinking neuron that employs a synergy between leaky integrate-and-fire function and transient electroluminescence. The MBN device efficiently transmits memristive neuron's state in the form of optical spikes. The MBN device pioneers an artificial neuron that features CMOS-compatibility, and low hardware cost (no peripheral readout circuit, no electrical interconnect). This approach gives inspiration for an advanced 3D neuromorphic computing hardware paradigm. The photonically-linked SNN discussed here, and already achieving satisfactory results on audio and image classification tasks, demonstrates the MBN's potential as a promising candidate for constructing three-dimensional interconnected electro-optical neural network hardware.

Methods

Fabrication of Memristive Blinking Neuron.

The memristive blinking neuron devices are fabricated on a glass substrate with low autofluorescence. An electrode metal of 30/15 nm Ag/Au is initially deposited. The fine-featured antenna is patterned using electron beam lithography, followed by ion milling to form the antenna structure. Subsequently, the contact pad is then deposited and patterned using photolithography. A 150 nm PMMA top dielectric layer is then spin-coated and annealed at 180°C as the switching medium.

Experimental Measurement.

The configuration for optical assessments and signal processing is illustrated in Fig. S7. Our setup features an inverted microscope (ECLIPSE, Nikon) as the sample stage. Emitted photons in the z^- plane are captured using an oil immersion objective (100x, 1.49 NA, Nikon) for high-resolution imaging. The photons are then directed through a selected output port to either an avalanche photodiode (SPCM-780-14-BR1, Excelitas) or a CMOS camera (C15550-20UP, HAMAMATSU) to capture temporal and spatial features, respectively. An arbitrary pulse is generated by a Lock-in Amplifier (UHFLI, Zurich Instruments), which applies voltage signals directly to the sample's surface via GGB DC probes. Fig. S7(a) illustrates the setup for the current measurement, while Fig. S7(b) and S7(c) shows the setup for photonic response measurement.

Spiking Neural Network implementation.

In this work, we construct two spiking neural networks for keyword spotting and image classification tasks, respectively. We preprocess the voice and image signals and encode them into temporal spike sequences with a layer of spiking neurons, more details are discussed in Supplementary Information Section IX and X. The network for KWS consists of a convolutional preprocessing module and two fully connected layers, while the network for image classification consists of three spiking convolutional layers and a classifier. In the hardware implementation, the synaptic weights are represented by different voltage amplitudes on the neuron, and the spike output is generated by binarizing the photon count produced by the corresponding MBN device. The neuron with the highest firing rate in the output layer determines the network's classification class.

References

1 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436-444, doi:10.1038/nature14539 (2015).

2 Reuther, A. *et al.* in *2022 IEEE High Performance Extreme Computing Conference (HPEC)*. 1-10 (IEEE).

3 Dhar, P. The carbon impact of artificial intelligence. *Nature Machine Intelligence* **2**, 423-425, doi:10.1038/s42256-020-0219-9 (2020).

4 Baldo, M. *Introduction to nanoelectronics*. (MIT OpenCourseWare, 2011).

5 Moore, M. J. A. J. International roadmap for devices and systems. (2020).

6 Samsi, S. *et al.* in *2023 IEEE High Performance Extreme Computing Conference (HPEC)*. 1-9 (IEEE).

7 Taylor, M. B. J. I. M. A landscape of the new dark silicon design regime. **33**, 8-19 (2013).

8 Kim, J. *et al.* in *Proceedings of the 34th annual international symposium on Computer architecture*. 138-149.

9 Esmaeilzadeh, H., Blem, E., St. Amant, R., Sankaralingam, K. & Burger, D. in *Proceedings of the 38th annual international symposium on Computer architecture*. 365-376.

10 Young, N., Collins, C. & Kaas, J. Cell and neuron densities in the primary motor cortex of primates. **7**, doi:10.3389/fncir.2013.00030 (2013).

11 Christensen, D. V. *et al.* 2022 roadmap on neuromorphic computing and engineering. *Neuromorphic Computing and Engineering* **2**, 022501, doi:10.1088/2634-4386/ac4a83 (2022).

12 Choi, S., Yang, J. & Wang, G. Emerging Memristive Artificial Synapses and Neurons for Energy-Efficient Neuromorphic Computing. *Advanced Materials* **32**, 2004659, doi:https://doi.org/10.1002/adma.202004659 (2020).

13 Gökgöz, B., Gül, F. & Aydın, T. An overview memristor based hardware accelerators for deep neural network. *Concurrency and Computation: Practice and Experience* **36**, e7997, doi:https://doi.org/10.1002/cpe.7997 (2024).

14 Goossens, A. S. *et al.* Memristive Memory Enhancement by Device Miniaturization for Neuromorphic Computing. *Advanced Electronic Materials* **9**, 2201111, doi:https://doi.org/10.1002/aelm.202201111 (2023).

15 Terabe, K., Hasegawa, T., Nakayama, T. & Aono, M. Quantized conductance atomic switch. *Nature* **433**, 47-50, doi:10.1038/nature03190 (2005).

16 Valov, I. *et al.* Nanobatteries in redox-based resistive switches require extension of memristor theory. *Nature Communications* **4**, 1771, doi:10.1038/ncomms2784 (2013).

17 Akinaga, H. & Shima, H. Resistive Random Access Memory (ReRAM) Based on Metal Oxides. *Proceedings of the IEEE* **98**, 2237-2251, doi:10.1109/JPROC.2010.2070830 (2010).

18 Jo, S. H. *et al.* Nanoscale Memristor Device as Synapse in Neuromorphic Systems. *Nano Letters* **10**, 1297-1301, doi:10.1021/nl904092h (2010).

19 Le Gallo, M. *et al.* A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference. *Nature Electronics* **6**, 680-693, doi:10.1038/s41928-023-01010-1 (2023).

20 Wang, Z. *et al.* Fully memristive neural networks for pattern classification with unsupervised learning. *Nature Electronics* **1**, 137-145, doi:10.1038/s41928-018-0023-2 (2018).

21 Aamir, S. A., Müller, P., Hartel, A., Schemmel, J. & Meier, K. in *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference*. 71-74.

22 Qiao, N. *et al.* A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. *Frontiers in neuroscience* **9**, doi:10.3389/fnins.2015.00141 (2015).

23 Sun, W. *et al.* Understanding memristive switching via in situ characterization and device modeling. *Nature Communications* **10**, 3453, doi:10.1038/s41467-019-11411-6 (2019).

24 Yang, R., Huang, H.-M. & Guo, X. Memristive Synapses and Neurons for Bioinspired Computing. *Advanced Electronic Materials* **5**, 1900287, doi:https://doi.org/10.1002/aelm.201900287 (2019).

25 Corinto, F. & Forti, M. Memristor Circuits: Flux—Charge Analysis Method. *IEEE Transactions on Circuits and Systems I: Regular Papers* **63**, 1997-2009, doi:10.1109/TCSI.2016.2590948 (2016).
26 Amirsoleimani, A. *et al.* In-Memory Vector-Matrix Multiplication in Monolithic Complementary Metal–Oxide–Semiconductor-Memristor Integrated Circuits: Design Choices, Challenges, and Perspectives. *Advanced Intelligent Systems* **2**, 2000115, doi:https://doi.org/10.1002/aisy.202000115 (2020).
27 Güngördü, A. D., Dündar, G. & Yelten, M. B. in *2020 IEEE International Symposium on Circuits and Systems (ISCAS).* 1-5.
28 Aguirre, F. *et al.* Hardware implementation of memristor-based artificial neural networks. *Nature Communications* **15**, 1974, doi:10.1038/s41467-024-45670-9 (2024).
29 Zhou, Z. *et al.* Prospects and applications of on-chip lasers. **3**, 1 (2023).
30 Shastri, B. J. *et al.* Photonics for artificial intelligence and neuromorphic computing. *Nature Photonics* **15**, 102-114, doi:10.1038/s41566-020-00754-y (2021).
31 Lindenmann, N. *et al.* Photonic wire bonding: a novel concept for chip-scale interconnects. *Opt. Express* **20**, 17667-17677, doi:10.1364/OE.20.017667 (2012).
32 Hamerly, R., Bernstein, L., Sludds, A., Soljačić, M. & Englund, D. Large-Scale Optical Neural Networks Based on Photoelectric Multiplication. *Physical Review X* **9**, 021032, doi:10.1103/PhysRevX.9.021032 (2019).
33 Zhou, T. *et al.* Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nature Photonics* **15**, 367-373, doi:10.1038/s41566-021-00796-w (2021).
34 Ai, C. Y. *et al.* in *Proceedings of the Int'l Image Sensor Workshop (IISW), Crieff, UK.* 22-25.
35 Uchiyama, M. *et al.* in *International Image Sensor Workshop (IISW).*
36 Hughes, R. W. & Warner, M. LEDs driven by AC without transformers or rectifiers. *Scientific Reports* **11**, 963, doi:10.1038/s41598-020-80617-2 (2021).
37 Żak, M. *et al.* Bidirectional light-emitting diode as a visible light source driven by alternating current. *Nature Communications* **14**, 7562, doi:10.1038/s41467-023-43335-7 (2023).
38 Li, X. *et al.* Bright semiconductor single-photon sources pumped by heterogeneously integrated micropillar lasers with electrical injections. *Light: Science & Applications* **12**, 65, doi:10.1038/s41377-023-01110-9 (2023).
39 Qian, H. *et al.* Efficient light generation from enhanced inelastic electron tunnelling. *Nature Photonics* **12**, 485-488, doi:10.1038/s41566-018-0216-2 (2018).
40 Cheng, B. *et al.* Atomic scale memristive photon source. *Light: Science & Applications* **11**, 78, doi:10.1038/s41377-022-00766-z (2022).
41 Malchow, K. *et al.* Self-Induced Light Emission in Solid-State Memristors Replicates Neuronal Biophotons. *ACS Nano* **18**, 24004-24011, doi:10.1021/acsnano.4c02924 (2024).
42 Cao, R. *et al.* Compact artificial neuron based on anti-ferroelectric transistor. *Nature Communications* **13**, 7018, doi:10.1038/s41467-022-34774-9 (2022).
43 Seok Jeong, D., Kim, I., Ziegler, M. & Kohlstedt, H. Towards artificial neurons and synapses: a materials point of view. *RSC Advances* **3**, 3169-3183, doi:10.1039/C2RA22507G (2013).
44 Atluri, P. P. & Regehr, W. G. Determinants of the Time Course of Facilitation at the Granule Cell to Purkinje Cell Synapse. *The Journal of Neuroscience* **16**, 5661, doi:10.1523/JNEUROSCI.16-18-05661.1996 (1996).
45 Eshraghian, J. K., Wang, X. & Lu, W. D. Memristor-Based Binarized Spiking Neural Networks: Challenges and applications. *IEEE Nanotechnology Magazine* **16**, 14-23, doi:10.1109/MNANO.2022.3141443 (2022).
46 Li, X., Zhang, G., Huang, H. H., Wang, Z. & Zheng, W. in *2016 45th International Conference on Parallel Processing (ICPP).* 67-76.
47 Wong, C. Cubic millimetre of brain mapped in spectacular detail. *Nature* **629**, 739-740, doi:10.1038/d41586-024-01387-9 (2024).
48 Besrour, M. *et al.* in *2022 20th IEEE Interregional NEWCAS Conference (NEWCAS).* 148-152 (IEEE).
49 Moradi, S., Bhave, S. A. & Manohar, R. in *2017 IEEE Symposium Series on Computational Intelligence (SSCI).* 1-5 (IEEE).
50 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, doi:10.1038/s41586-019-1157-8 (2019).
51 Wang, T. *et al.* Image sensing with multilayer nonlinear optical neural networks. *Nature Photonics* **17**, 408-415, doi:10.1038/s41566-023-01170-8 (2023).

52 Chen, Y. *et al.* All-analog photoelectronic chip for high-speed vision tasks. *Nature* **623**, 48-57, doi:10.1038/s41586-023-06558-8 (2023).
53 Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501-506, doi:10.1038/s41586-022-04714-0 (2022).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIMBNsubmission.docx