Opto-Electronic Advances

Article

November 2023, Vol. 6, No. 11

DOI: 10.29026/oea.2023.230140

Photonic integrated neuro-synaptic core for convolutional spiking neural network

Shuiying Xiang[1]*, Yuechun Shi[2]*, Yahui Zhang[1], Xingxing Guo[1], Ling Zheng[3], Yanan Han[1], Yuna Zhang[1], Ziwei Song[1], Dianzhuang Zheng[1], Tao Zhang[1], Hailing Wang[4], Xiaojun Zhu[5], Xiangfei Chen[6], Min Qiu[7], Yichen Shen[8], Wanhua Zheng[4] and Yue Hao[1]

Neuromorphic photonic computing has emerged as a competitive computing paradigm to overcome the bottlenecks of the von-Neumann architecture. Linear weighting and nonlinear spike activation are two fundamental functions of a photonic spiking neural network (PSNN). However, they are separately implemented with different photonic materials and devices, hindering the large-scale integration of PSNN. Here, we propose, fabricate and experimentally demonstrate a photonic neuro-synaptic chip enabling the simultaneous implementation of linear weighting and nonlinear spike activation based on a distributed feedback (DFB) laser with a saturable absorber (DFB-SA). A prototypical system is experimentally constructed to demonstrate the parallel weighted function and nonlinear spike activation. Furthermore, a four-channel DFB-SA laser array is fabricated for realizing matrix convolution of a spiking convolutional neural network, achieving a recognition accuracy of 87% for the MNIST dataset. The fabricated neuro-synaptic chip offers a fundamental building block to construct the large-scale integrated PSNN chip.

Keywords: neuromorphic computation; photonic spiking neuron; photonic integrated DFB-SA array; convolutional spiking neural network

Xiang SY, Shi YC, Zhang YH, Guo XX, Zheng L et al. Photonic integrated neuro-synaptic core for convolutional spiking neural network. *Opto-Electron Adv* **6**, 230140 (2023).

Introduction

The deep neural network has developed rapidly and achieved record-breaking performance in a broad range of applications, including computer vision, natural language processing, and other fields. These applications produce huge amounts of data that need to be processed, which calls for advanced processors with high speed, high throughput and low latency. While Moore's Law approaches saturation and conventional digital computers face the von-Neumann bottleneck due to the sep-

[1]State Key Laboratory of Integrated Service Networks, State Key Discipline Laboratory of Wide Bandgap Semiconductor Technology, Xidian University, Xi'an 710071, China; [2]Yongjiang laboratory, No. 1792 Cihai South Road, Ningbo 315202, China; [3]The School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; [4]Laboratory of Solid-State Optoelectronics Information Technology, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China; [5]School of Information Science and Technology, Nantong University, Nantong 226019, China; [6]The College of Engineering and Applied Sciences, Nanjing University, Nanjing 210023, China; [7]Key Laboratory of 3D Micro/Nano Fabrication and Characterization of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310024, China; [8]Lightelligence Group, Hangzhou 311121, China.

*Correspondence: SY Xiang, E-mail: syxiang@xidian.edu.cn; YC Shi, E-mail: yuechun-shi@ylab.ac.cn

Received: 9 August 2023; Accepted: 9 October 2023; Published online: 15 November 2023

aration of memory and processor units. Neuromorphic computing, which emulates the structure and mechanism of the brain, has emerged as a competitive computation paradigm in the post-Moore era[1–3]. Although the electronic neuromorphic computing systems based on the well-known complementary-metal-oxide-semiconductor (CMOS) technique or emerging electronic devices have made notable progress, they still face limitations in processing speed and energy efficiency[4–11]. As a promising alternative, photonic computing has garnered significant attention as it offers inherent advantages such as ultra-high speed, large bandwidth and massive parallelism. Different photonic computing approaches have been proposed due to their potentials for addressing the limitations of electronic counterparts[12–21].

In addition to the advancements in neuromorphic hardware, novel neural network models have also attracted significant attention. In contrast to traditional non-spiking artificial neural networks (ANNs), spiking neural networks (SNNs) have been proposed and designed to enable low power consumption computing on neuromorphic hardware by using discrete spike signals[22–24]. The implementation of SNNs on a photonic neuromorphic hardware holds great promise for a wide range of applications in latency-critical and power-constrained scenarios, such as autonomous driving and edge computing[25–29].

Spiking neurons serve as the fundamental units of SNNs, and are connected through plastic synapses. In an SNN, the spiking neuron performs the nonlinear spike activation, while the synapse performs the linear weighting function. For the majority of reported photonic neural networks, only linear operation was realized in the photonics domain[29–38]. For instance, two mainstream approaches, namely the coherent synaptic network based on the Mach-Zehnder interferometer (MZI)[30,32,35–37] and the incoherent synaptic work based on the microring resonator (MRR)[29,38] have been widely explored due to their compatibility with CMOS-compatible silicon photonics platform. However, these approaches suffer from serious loss issues and are not well-suited for implementing nonlinear computations directly in the photonic domain. In the reported photonic neural network architectures, nonlinear computations are mainly implemented electronically, relying on high-speed photodetectors (PDs) and analog-to-digital (AD) converters to convert the optical linear computation results back to the digital domain. Such hybrid architectures, consisting of photonic synapses and electronic spiking neurons, present obstacles for the photonic implementation of multi-layer or deep neural networks. The frequent optic-electro (OE) and electro-optic (EO) conversions, as well as the AD and digital-to-analog (DA) conversions, hinder the seamless integration of photonic functionalities across multiple layers.

Tremendous efforts have been made to realize photonic spiking neurons, aiming to enable all-optical SNNs without the need for frequent OE/EO and AD/DA conversions. Among various simplified spiking neuron models in computational neuroscience, the leaky integrate-and-fire (LIF) model has gained popularity due to its simplicity. Optical implementations of the LIF neuron have received considerable attention[39]. Several optical implementations of the LIF neuron have been explored, including polarization switching vertical-cavity surface-emitting lasers (VCSELs)[40,41], VCSEL with a saturable absorber (VCSEL-SA)[26,42], graphene excitable laser[43], hybrid integrated phase-change material (PCM) and MRR[25,44], micropillar lasers[45], integrated distributed feedback (DFB) laser and PD[46], passive microresonator[47], and integrated Fabry–Perot laser with a saturable absorber (FP-SA)[28,48]. However, these photonic spiking neurons have been designed and fabricated independently from the photonic synapses, limiting the scalability of fully-functional photonic SNNs.

Here, we proposed, fabricated and demonstrated a photonic neuro-synaptic chip that enables simultaneous implementation of both linear weighting and nonlinear spike activation based on a photonic integrated DFB laser with an intracavity saturable absorber (DFB-SA) for the first time. By precisely tuning the external optical injection and the time-varying bias current of the gain region of the DFB-SA laser, we experimentally demonstrate both the intrinsic excitability plasticity and synaptic plasticity in the same DFB-SA laser. The nonlinear spike activation and linear weighting can be simultaneously realized in a single DFB-SA laser, which is the main advantage of the proposed photonic neuro-synaptic chip. In addition, parallel linear weighting and nonlinear spike activation were demonstrated in a constructed prototypical neuromorphic photonics system consisting of two pre-synaptic DFB-SA lasers and one post-synaptic FP-SA laser. To demonstrate the scalability of the DFB-SA laser, a four-channel photonic neuro-synaptic array was further fabricated and applied to perform dot product between two vectors. Through time-multiplexing matrix convolution, we successfully

demonstrated pattern classification tasks. We further developed a theoretical model to numerically simulate the photonic integrated DFB-SA laser, validating our experimental findings. The integration of linear weighting and nonlinear spike activation in a single chip opens up new possibilities for efficient and scalable photonic neuromorphic computing, presenting a promising building block for constructing multi-layer photonic SNN hardware within the Indium Phosphide (InP) integration platform.

Experimental setup and Method

Principle for photonic neuro-synaptic core

In a biological neural network, the neurons are connected by plastic synapses. As depicted in Fig. 1(a), the dendrites receive external stimuli or pre-synaptic spikes, while the soma converts these inputs into spike events and performs nonlinear spike activation. The synapses adjust the connection strength, commonly referred to as weight in ANNs. The weighted spike is then transmitted to the dendrites of the neighboring neurons. To address the challenge of implementing both linear weighting and nonlinear spike activation in a single device, we proposed, designed and fabricated a photonic integrated DFB-SA laser, which we refer to as a photonic neuro-synaptic chip. The structure of the DFB-SA laser chip is illustrated in Fig. 1(b). We introduce a SA region in the resonant cavity of the DFB laser to obtain Q-switching effect. The grating is designed with a sampled grating. We shift the half period of the sampling structure in the middle of the gain region of the DFB laser, which can equivalently introduce a π phase shift (π-EPS). This grating structure can be fabricated with the reconstruction-equivalent-chirp (REC) technique, which allows for a large-scale DFB laser array with high wavelength precision[49]. Additionally, anti-reflection (AR) and high-reflection (HR) coating are applied to the two laser facets to enhance the light emission power, and the SA region is positioned near the HR side. Figure 1(c) provides an overview of the spike processing principle in the DFB-SA laser. The gain region of the DFB-SA laser is driven by a current source, denoted as the gain current I_G, while the SA region is reversely driven by a voltage source, denoted as V_{SA}. By applying a time-varying current to the gain region, we achieve dynamically linear weighting functionality. The interaction between photons and electrons in the gain and SA regions enables the emulation of a LIF-type spiking neuron.

We further designed and fabricated a four-channel photonic neuro-synaptic array to implement two-dimensional (2D) convolution for spiking convolutional neural networks (CNNs). To fully leverage the temporal dynamics of the SNN, we have demonstrated time-multiplexing matrix convolution. The fabricated four-channel DFB-SA laser array and the compact packaged module are depicted in Fig. 1(d). The principle of using the DFB-SA laser array to achieve spike-based matrix convolution is illustrated in Fig. 1(e). The input is electro-optically modulated and optically injected into the DFB-SA laser array, while the weight is directly imported to the bias current of the gain region. Thus, the DFB-SA laser array performs dot product operations and acts as a photonic dot product core. The input and weight signals were generated using an FPGA (Zynq UltraScale+ RFSoC ZCU216) equipped with a high-speed AD/DA array. The FPGA was controlled by a digital computer. To generate optical carriers, four-channel continuous-wave tunable laser sources were utilized. The input signals were modulated using the Mach-Zehnder modulators (MZMs). Polarization controllers were employed to align the polarization state. The modulated outputs of the MZMs were then optically injected into the DFB-SA laser array via three-port optical circulators. The kernel signals were applied to the gain current of the DFB-SA lasers array via bias tees. The outputs of four DFB-SA lasers were combined by a four-port optical coupler. The combined signal was detected by a photodetector (PD, Agilent/HP 11982A) and then recorded by an oscilloscope (OSC, Keysight DSOV334A).

Model of the DFB-SA laser

In order to gain a deeper understanding of the intrinsic excitability plasticity and synaptic plasticity in the DFB-SA laser, we have developed a comprehensive model of the DFB-SA laser based on the time-dependent coupled-wave equations. We modified the model to incorporate the gain region, SA region, and an external optical injection term[50]. The rate equations governing the carrier density in both the gain and absorber regions, and the coupled-mode equations describing the behavior of the forward and backward traveling waves, are detailed in the Supplementary information Section 1.

Results

Intrinsic neuron excitability plasticity

The experimental setup for emulating the neuronal in-

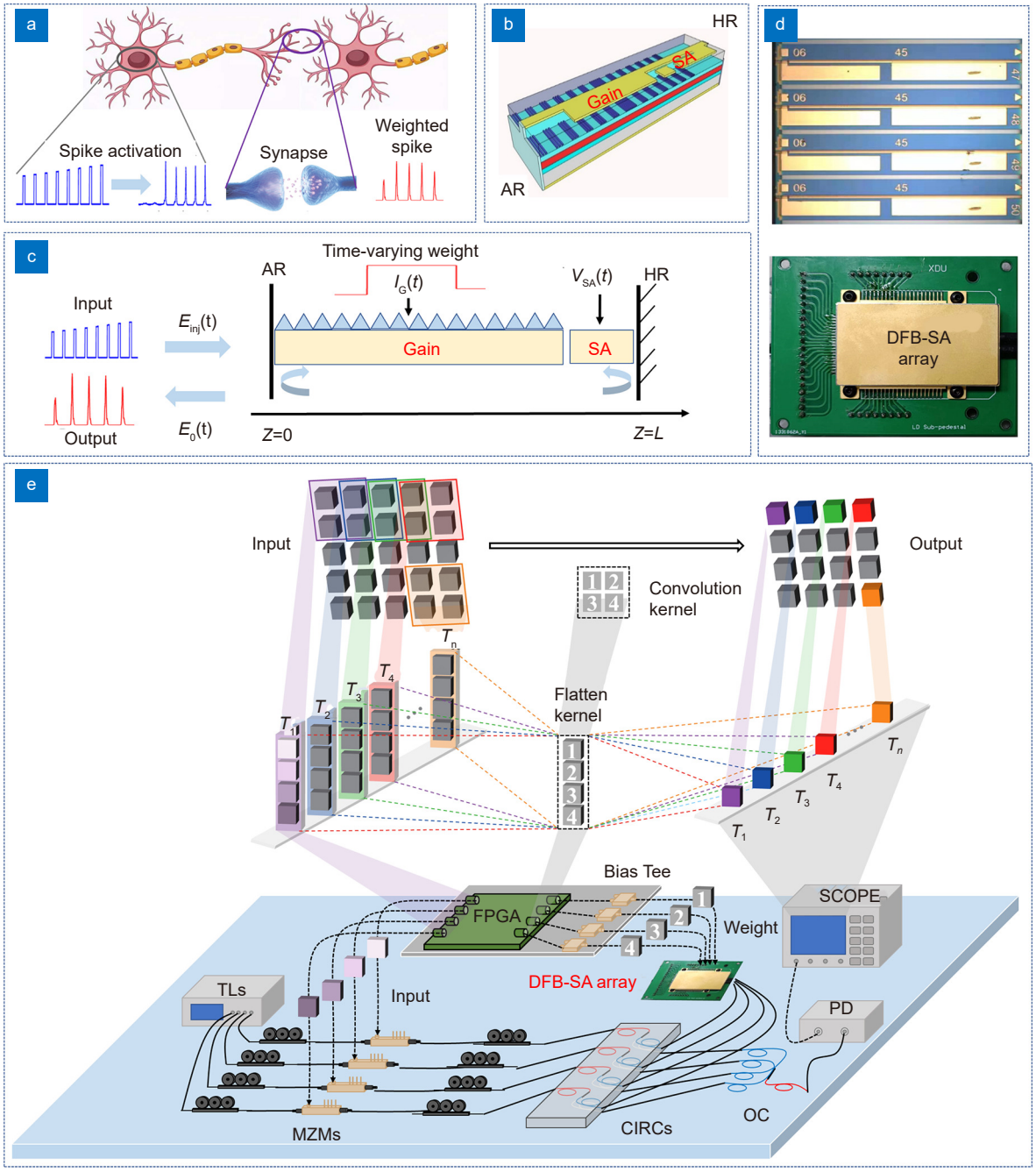

Fig. 1 | The operation principle of photonic neuron-synaptic core. (**a**) The schematic diagram of biological neuron. (**b**) The structure and micrograph picture of the designed DFB-SA laser chip. (**c**) The spike processing in a DFB-SA laser. (**d**) The micrograph picture of the fabricated DFB-SA laser array chip and the compact packaged module. (**e**) Spike-based matrix convolution based on the four-channel DFB-SA laser array.

trinsic plasticity and linear weighting using a DFB-SA laser is presented in Fig. 2(a). An arbitrary waveform generator (AWG) (Tektronix AWG70001A) was used to generate the electronic stimulus and the time-varying modulated current for the DFB-SA laser. The time-varying current was combined with a direct current source using a bias tee and then applied to the gain region of DFB-SA laser. A tunable laser (TL, AQ2200-136 TLS module) provided the optical carrier. The electro-optic conversion was realized using an MZM. The modulated optical stimulus was then injected into the gain section of the DFB-SA laser via an optical circulator. In our experiment, we maintained a fixed temperature of 25 °C. The threshold current of the DFB-SA laser was measured to be approximately I_G=86 mA when V_{SA}=0 V and I_G=94 mA when V_{SA}= −0.4 V. The optical spectrum of a free-running DFB-SA laser is presented in Fig. 2(b). The side mode suppression ratio was found to be approximately

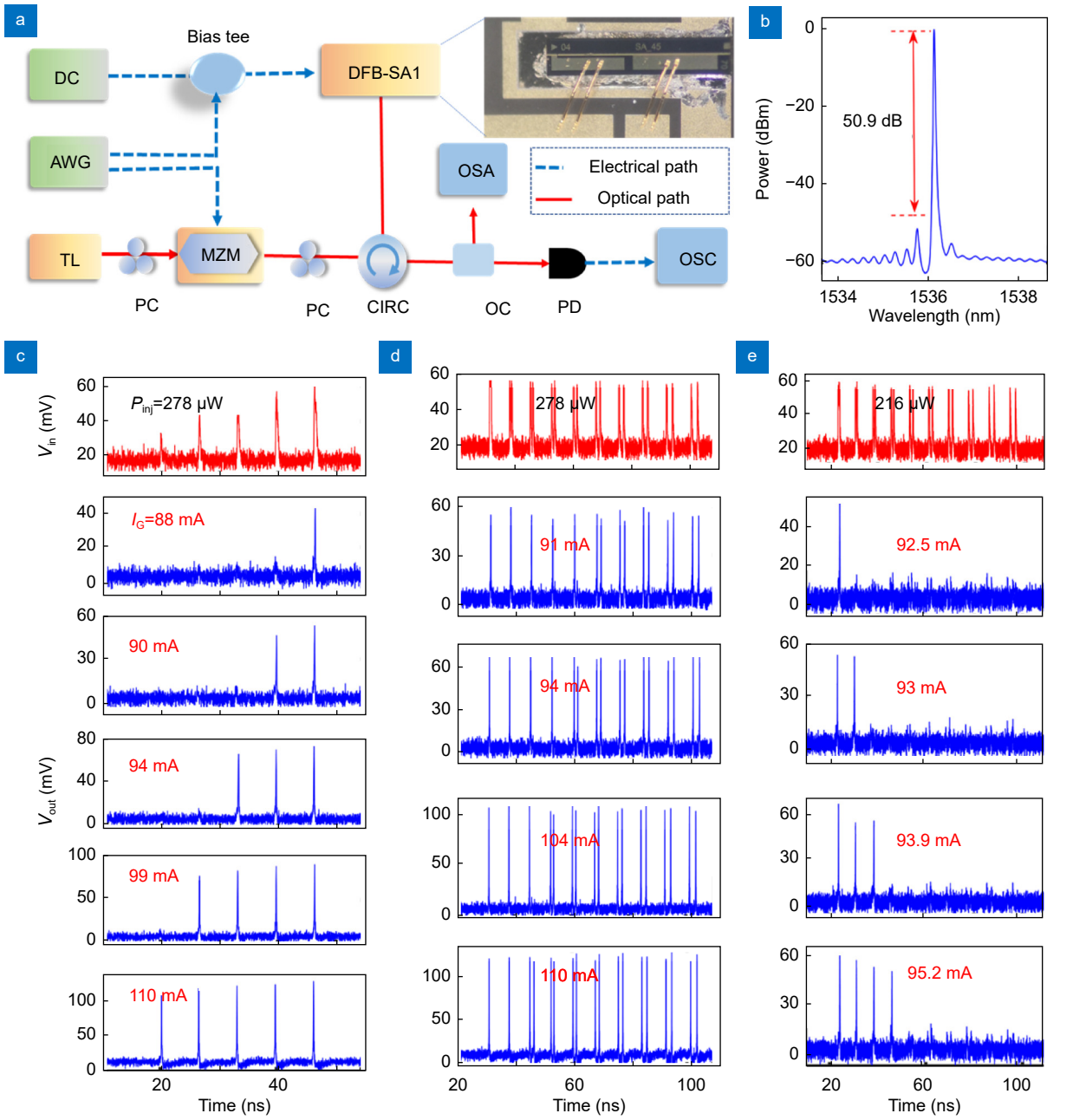

Fig. 2 | Experimental demonstration of intrinsic excitability plasticity in a single fabricated DFB-SA laser chip. (**a**) Experimental setup of a DFB-SA laser for emulating the neuronal intrinsic plasticity and linear weighting. (**b**) Optical spectrum of the free-running DFB-SA laser. (**c**) Different excitability threshold, (**d**) different refractory period, and (**e**) different temporal integration behavior under different gain currents. V_{SA}=−0.4V. The inset in (a) corresponds to the wire bonding of the fabricated DFB-SA laser chip.

50.9 dB, indicating a high level of suppression of unwanted side modes.

For a LIF spiking neuron, temporal integration, spike threshold and generation as well as the refractory period are crucial information processing mechanisms. Here, we successfully demonstrated these spike-based nonlinear processing mechanisms using the fabricated DFB-SA laser chip. As shown in Fig. 2(c), we designed five input pulses with varying amplitudes for a given external stimulus strength and injection power, and obtained different nonlinear spike activations simply by adjusting the gain current. For instance, when I_G=88 mA, only the stimulus pulse with the highest amplitude exceeded the excitability threshold, resulting in a successful response spike. With the increase of gain current, the number of responded spikes increased in a completely controllable manner. Note that, the external stimulus strength corresponding to the synapse weight remained fixed

throughout the experiment. The different spike response was solely achieved by adjusting the excitability threshold of the DFB-SA laser spiking neuron. This demonstrates the presence of intrinsic excitability threshold plasticity in the fabricated DFB-SA laser, which is similar to its biological counterpart[51]. Thus, nonlinear spike activation for a photonic SNN can be achieved in the DFB-SA laser.

Similarly, we fixed the external stimulus strength, and designed 10 pulse pairs with increasing inter-spike interval (ISI) to explore the intrinsic plasticity of the refractory period and temporal integration behavior. Here, the ISI for each input pulse pair is 0.4 ns, 0.6 ns, 0.8 ns, 1 ns, 1.2 ns, 1.4 ns, 1.6 ns, 1.8 ns, 2 ns, and 2.2 ns, respectively. As presented in Fig. 2(d), for different input pulse pairs, the number of response spikes was either 1 or 2 depending on the ISI. For I_G=91 mA, the first 5 pulse pairs with relatively small ISIs resulted in a single response spike, while the last 5 pulse pairs with larger ISIs triggered two response spikes each. With the increase of I_G, the number of pulse pairs leading to a single response spike gradually decreased to 4, 3 and 2, respectively. This behavior emulates the refractory period of a biological neuron, which means that the spiking neuron cannot produce another response spike to the stimulus in a short time due to the carrier recovery process, if it just responded a spike to the preceding stimulus[28]. Note, the spiking response rate is limited by the inverse of the refractory period. For I_G=91 mA, the refractory period is between 1.2 ns and 1.4 ns, corresponding to a spiking response rate of around 0.77 GHz (inverse of 1.3 ns). Similarly, the spiking response rate is about 0.91 GHz for I_G=94 mA, 1.11 GHz for I_G=104 mA, and 1.43 GHz for I_G=110 mA. Note, the refractory period is mainly determined by the carrier lifetime, which can be reduced by using a shorter cavity length. Thus, it is desirable to further optimize the DFB-SA laser chip to reduce the refractory period and increase the spike processing speed. To demonstrate the temporal integration behavior, the injecting power was decreased to a level where a single stimulus pulse alone could not elicit a response spike. Note, the excitability threshold is determined together by the injection power and bias current. As the injection power is reduced, the bias current needs to be adjusted accordingly in order to achieve spike response. In Fig. 2(e), when I_G=92.5 mA, only the first pulse pair with the smallest ISI could be temporally integrated, exceeding the excitability threshold and eliciting a response spike. With the increase of gain current, the number of response spikes also increased. In other words, the response spike resulting from the temporal integration effect could be controlled by simply adjusting the gain current of the DFB-SA laser.

Note, the demonstrated intrinsic neuron excitability plasticity implies that, when developing a supervised training algorithm for a photonic SNN consisting of the DFB-SA laser, not only the weights can be considered as adjustable parameters, but the excitability threshold of the DFB-SA laser can also be trainable to accelerate the training process. Thus, a novel hardware-aware training algorithm that combines both the weight and excitability threshold modulation is highly desirable to enhance the performance of photonic SNNs.

Continuously tunable synaptic plasticity

Next, we experimentally demonstrated the linear weighting function of the fabricated DFB-SA laser chip by employing a statically-varying bias current. We defined a periodic optical pulse train as the external stimulus and adjusted the gain current, ensuring that the DFB-SA laser operated in an LIF-like response manner. As illustrated in Fig. 3(a), the peak amplitude of the response spike gradually increased as the gain current was increased, ranging from 90 mA to 110 mA in 4 mA increments. In Fig. 3(b), we present long time traces of the weighted optical spike trains for I_G=90 mA, 100 mA and 110 mA. It is verified that a higher gain current contributes to a larger spike amplitude. Moreover, for a given gain current, the amplitude of the weighted spike maintains a relatively stable state across different periods, indicating the stability of the weighting function. We further present the peak amplitude of the response spike as a function of the gain current in Fig. 3(c). The bias current is varied from 88 mA to 112 mA. Here, the stimulus pulse with sufficiently large intensity is required to ensure that each can trigger a response spike in the DFB-SA laser for each bias current. Notably, the peak amplitude of the weighted spike increases almost linearly with the gain current, which can be attributed to the gain-switched operation mechanism. Consequently, in a photonic SNN, the synapse weight can be readily mapped to the gain current of the DFB-SA laser, facilitating straightforward weight modulation.

By simply tuning the bias current of the gain region of the DFB-SA laser, we successfully achieved precise control over the amplitude of the weighted optical spikes,

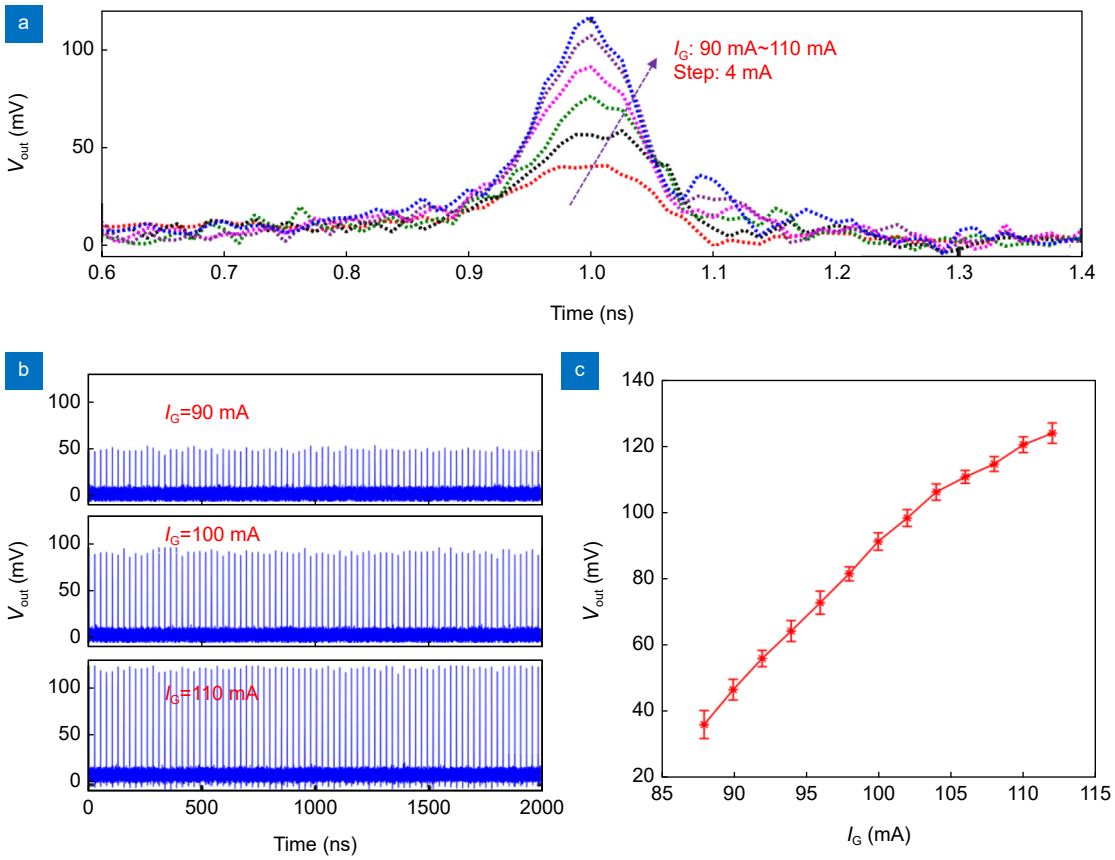

Fig. 3 | Continuously tunable weighted spike output for different static gain currents in a single fabricated DFB-SA laser chip. (**a**) The weighted output of a single spike for different gain currents. (**b**) The weighted spike trains for three representative gain current. (**c**) The spike peak amplitude as a function of gain current.

thus enabling a photonic integrated spiking neuron with incorporated spike weighting capability. Note, in comparison to biological counterparts that operate on millisecond timescales, the photonic neuro-synaptic unit utilizing the fabricated DFB-SA laser offers significantly faster operation speed in the sub-nanosecond range, thanks to the short carrier lifetime inherent to these chips. Therefore, the proposed photonic neuron-synaptic unit avoids the use of additional photonic weighting elements, making it highly desirable for the further scalability of photonic SNN hardware.

Simultaneous implementation of nonlinear spike activation and linear weighting

In the following, we considered the scenario where a dynamically time-varying gain current was applied to the DFB-SA laser, allowing for the simultaneous implementation of nonlinear spike activation and linear weighting. Note, the nonlinear spike activation is implemented due to the excitability threshold modulation. For the purpose of comparison, we employed a time-varying optical input stimulus and considered different constant gain currents, as depicted in Fig. 4(a). It is evident that, for a given gain current, as long as the stimulus intensity exceeds the excitability threshold, the spike amplitude remained nearly constant regardless of varying stimulus strengths. In addition, the excitability threshold was decreased with the increase of gain current, enabling a greater number of input pulses that can trigger response spikes. However, it is important to note that the spike amplitude differs across different gain currents, with larger gain currents resulting in higher spike amplitudes.

Then, we considered a fixed input strength and applied a dynamically modulated gain current. In this experiment, the repetition rate of the input pulse train was set as 0.5 GHz. As presented in Fig. 4(b), the gain current was modulated using 5 discrete constant levels with an AWG corresponding to 5 distinct weight values. The modulated current was then combined with a static bias current via a bias tee. The total gain current can be expressed as I_G=I_{G0}+I_M, where I_{G0} =80 mA denotes the static bias current and I_M represents the modulated bias

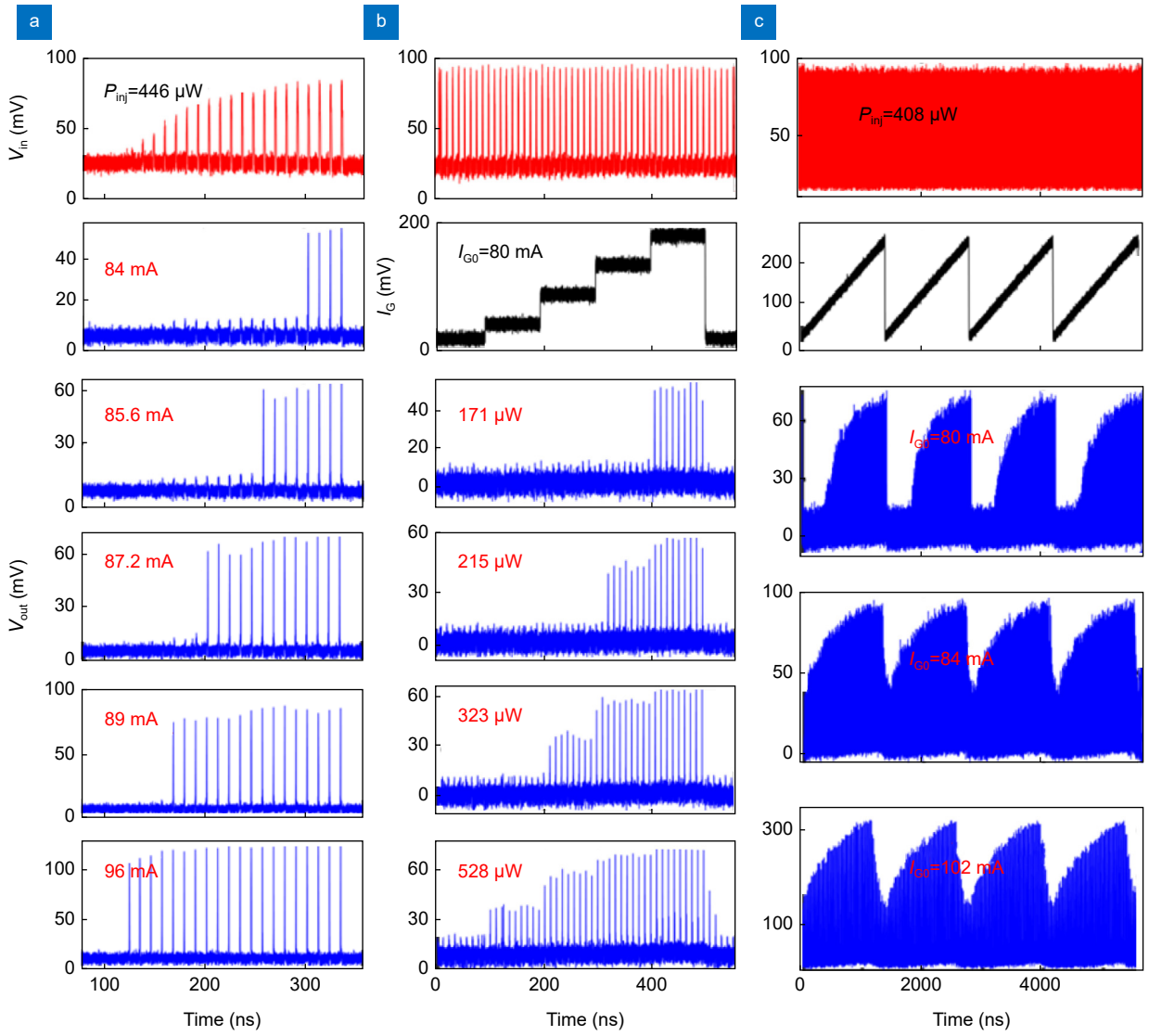

Fig. 4 | Experimental demonstration of simultaneous implementation of neuron excitability threshold and synaptic linear weighting functions in a single DFB-SA laser. (**a**) Corresponds to static gain current at different level. (**b**) and (**c**) Correspond to step-wise and continuously-varying gain current, respectively. In (b) I_{G0}=80 mA, $I_G = I_M + I_{G0}$.

current. It is evident that each case exhibited a different excitability threshold. When the injection power was 171 μw, only the highest level of I_M resulted in spike output. Increasing the injection power to 215 μW led to weighted spikes with two distinguishable amplitude levels. Namely, only the last two levels of I_M lead to spike activation. The activated spike amplitudes are modulated by different I_M corresponding to distinct weights. Further increasing the injection power enabled the attainment of three or four clusters of weighted spikes, each with varying spike amplitudes. In addition, we also considered the scenario of continuously time-varying gain current. As depicted in Fig. 4(c), the excitability threshold plasticity and linear weighting function was simultaneously achieved by simply modulating the gain current of the DFB-SA laser. Thus, the results demonstrated that the single DFB-SA laser chip enables simultaneous implementation of nonlinear spike activation and high-speed (0.5 GHz) linear weighting. The enlargement of a single period in Fig. 4(c) is further illustrated in the Supplementary information Section 2.

We also considered other input spike trains with higher repetition rates, and found that the linear weighting function was degraded. Some spikes cannot be normally weighted, which is mainly limited by the refractory period. Note, the spike weighting rate can be increased when a larger static bias current is utilized.

We further numerically reproduced the linear weighting function and nonlinear spike activation. The numerical results obtained from the model are presented in the

Supplement material Note 3. These results demonstrate that different gain currents result in distinct excitability thresholds. In addition, the modulation of gain current produces weighted spikes with distinguishable amplitude levels. Namely, the numerical results agree well with the experimental findings, thus providing a computational model that facilitates hardware-algorithm co-design and optimization of photonic SNNs utilizing the DFB-SA lasers.

Fully-functional photonic neuromorphic prototypical system

We further constructed a fully-functional prototypical photonic neuromorphic system utilizing the fabricated DFB-SAs for parallel linear weighting and nonlinear spike activation. Note, as the packaged DFB-SAs operate at different wavelengths, we adopted an FP-SA as the post-synaptic neuron because it can support wide bandwidth operation[28]. The system configuration is illustrated in Fig. 5(a), where DFB-SA1 and DFB-SA2 act as pre-synaptic photonic spiking neurons and the corresponding weighting devices, while the FP-SA serves as the post-synaptic photonic spiking neuron. The weighted spike outputs of the DFB-SAs were optically combined and injected into the FP-SA laser. The optical spectra of the DFB-SA lasers and FP-SA laser are displayed in Fig. 5(b). In this setup, the binary patterns 1101 and 0110 are employed as inputs for the DFB-SA1 and DFB-SA2, respectively. By setting different gain currents of the DFB-SAs, they performed digital-to-spike conversion and parallel linear weighting. The optical coupler then performed the summation of the two weighted spikes. The weighted sum was subsequently injected into the FP-SA for nonlinear spike activation.

The output of the FP-SA reveals that it can yield either 0100 or 1111, depending on the weighting conditions of the DFB-SAs or the excitability threshold of the FP-SA. Consequently, the constructed neuromorphic prototypical system can be flexibly reconfigured to execute spike-based AND as well as OR operations. Note, both the weighted addition operation and nonlinear spike activation are achieved within this photonic neuromorphic prototypical system based on the DFB-SAs and FP-SA, which can be integrated onto the same InP integration platform.

Time multiplexing matrix convolution with four-channel DFB-SA laser array

To demonstrate the scalability of the proposed photonic neuro-synaptic unit based on the DFB-SA, we fabricated a four-channel DFB-SA array and applied it to perform matrix convolution for a convolutional SNN. The threshold for each DFB-SA laser is about 86 mA when V_{SA}=0 V. The measured optical spectra for the four-channel DFB-SA laser array is presented in Supplementary information Section 5. The operation principle is illustrated in Fig. 6(a). In this configuration, the convolutional kernel size is 2×2, the stride step is 1. Each element in the convolutional kernel is mapped to a respective DFB-SA within the array. The gain currents of the fabricated DFB-SAs array are configured according to the trained kernels. The convolved results are obtained as the output of the optical coupler. The layout of the DFB-SA array is shown in the inset of Fig. 6(a), which exemplifies the potential for scalability of the photonic neuro-synaptic unit.

The spiking CNN was trained using a digital computer to classify the MNIST dataset[52], which consists of handwritten digits ranging from 0 to 9. The training set contains 60000 samples and the test set contains 10000 samples. The size of each image is 28×28 pixels. The spiking CNN consists of a convolutional layer with 16 convolutional kernels, an activation layer, a pooling layer, and two fully-connected layers, as depicted in Fig. 6(b). The trained 16 convolutional kernels were presented in Supplementary information Section 4. In this experiment, a hardware-software collaborative approach was adopted for inference. The photonic hardware was utilized to realize the convolutional layer. For evaluation, 100 test images were considered, and the experimental results for some representative input samples are presented in Fig. 6(c) (Please refer to Supplementary information Section 6 for additional input samples). The 16 convolutional kernels were time-multiplexed to take full advantage of the dynamically weight modulation capability of the DFB-SA laser array. The sampling rate of the DA was set at 6.4 Gbits/s. Considering the refractory period, 15 zeros were inserted between two pixels to match the FPGA bitwidth, resulting in a temporal duration of each pixel of 1/6.4 Gbits/s×16=2.5 ns. For each image, the temporal output of the convolved results for the 16 convolutional kernels includes 27×27×16 pixels. Thus, the total length of each measured output time series is 2.5 ns×27×27×16=29.16 μs. To demonstrate fast dynamically weight modulation, we calculated the convolved results row by row. Each row of pixels was convolved with

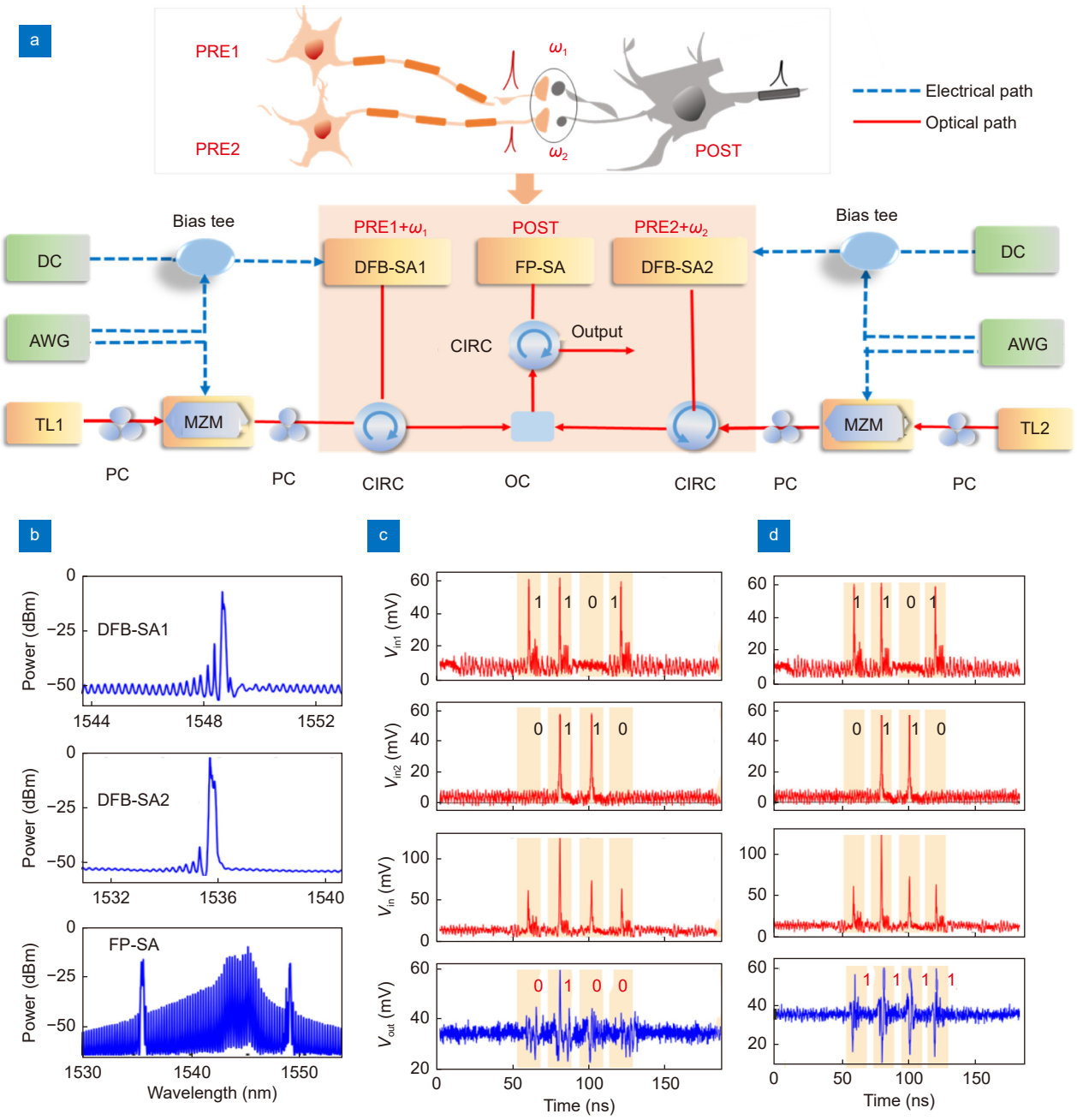

Fig. 5 | Spike-based neural computation in a photonic SNN consisting of two PRE DFB-SA lasers and one POST FP-SA laser. (**a**) Experimental setup of a full-functional prototypical photonic neuromorphic system. (**b**) Optical spectra of the DFB-SA lasers and FP-SA laser. (**c**) Spike-based AND. (**d**) Spike-based OR. With I_G=100.1 mA and V_{SA}=−0.1 V for the DFB-SA1 laser; I_G=102.1 mA and V_{SA}=−0.27 V for the DFB-SA2 laser; I_G=56.9 mA and V_{SA}=−0.25 V for the FP-SA laser.

16 different weights corresponding to different convolutional kernels. As a result, the temporal duration of a fixed weight was 2.5 ns×27=67.5 ns, yielding a modulation rate of approximately 14.8 MHz. Post-processing was performed with a digital computer. The feature maps are displayed in Supplementary information Section 7. After subsequent processing steps, the results indicate that pattern classification with a recognition accuracy of 87% was achieved. Note, with the same network structure, by using pure software inference, the recognition accuracy is 92.29% (for 10000 test images) and 90% (for 100 test images). The training accuracy, confusion matrix for software inference and hardware-software inference can be found in Fig. 6(d–f). The 3% loss in inference accuracy observed in the hardware-software inference scenario may be attributed to noise present in the experimental setup.

We further conducted numerical simulations to ex-

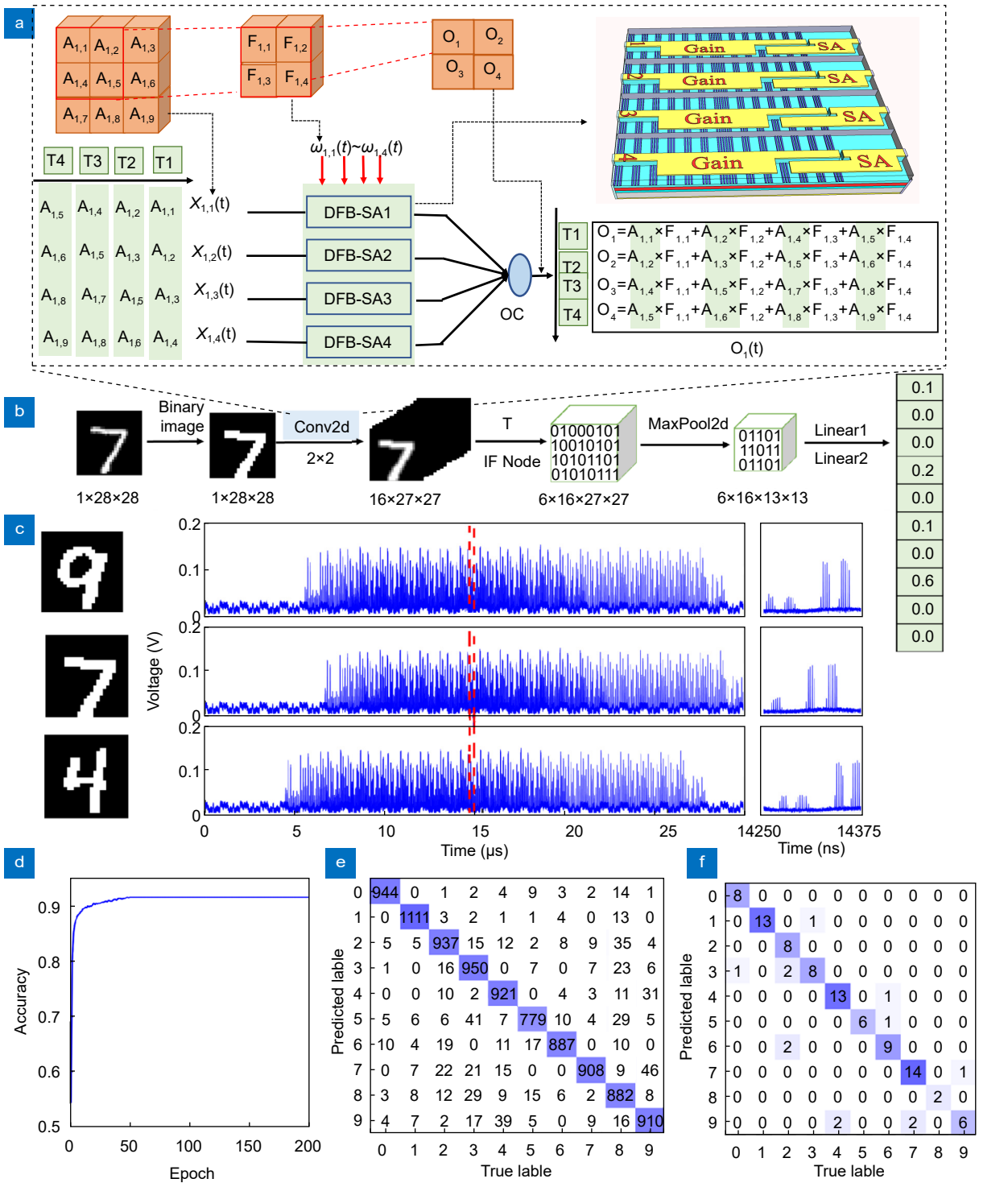

Fig. 6 | Experimental results of matrix multiplication with a four-channel DFB-SA laser array. (**a**) Schematic diagram of the parallel linear weight with the DFB-SA laser array. (**b**) Network structure for spiking CNN. (**c**) The temporal output of the convolved results for 16 convolutional kernels for three input samples, the right column represents the enlargement of the region denoted by the dashed lines. (**d**–**f**) Represent the training accuracy, confusion matrix for software inference and hardware-software inference.

plore time-multiplexing parallel dot product using our developed time-dependent traveling wave model. The numerical results of four-channel parallel linear weighting are presented in Supplementary information Section 8. The simulations revealed that the linear weighting operation could be effectively realized in parallel using this configuration. Moreover, we considered a larger convolution kernel size and simulated the case of a 3×3 convo-

lution kernel. The numerical results of 9-channel parallel linear weighting are provided in Supplementary information Section 9. It indicates that, 9-channel parallel dot product could be achieved by employing an array of nine DFB-SAs as the photonic dot product core. We further consider a spiking CNN network similar to the one shown in Fig. 6(b) but with 3×3 convolution kernels. The simulation results are presented in Supplementary information Section 10. The inference accuracy is 92.45% for the spiking CNN comprises one convolutional layer. Additionally, spiking CNN networks with two convolutional layers are also simulated, and the inference accuracy is 93.76% and 94.42% for convolution kernel size of 2×2 and 3×3, respectively. The numerical findings validate the feasibility and scalability of the proposed photonic neuro-synaptic unit for larger convolutional kernel operations.

Discussions and conclusions

The computation speed of the linear weighting operation in the fabricated four-channel DFB-SA laser array, with statically configured weights, can be estimated at 2×4×10 G=80 GOPS. On the other hand, the nonlinear computation speed for a single channel is approximately 2 G Spike/s, limited by the refractory period. The energy consumption per spike is approximately 19.99 pJ. The area occupied by a single DFB-SA laser chip is around 1500 μm ×300 μm=0.45 mm^2. It is important to note that there is significant potential for improving these metrics. Further optimization can be achieved by reducing the threshold of the DFB-SA laser to around or below 5 mA, and by reducing the area of a single DFB-SA laser chip to 300 μm×127 μm. The reconfigurable rate, which is associated with the speed of dynamics weight update, is estimated at 0.5 GHz, significantly faster than that of the thermos-optic phase shifters based on silicon photonics. Moreover, the dynamical weighting speed can be further increased to approximately 10 GHz by reducing the cavity length of the DFB-SA laser to 300 μm, which is compatible with the time-multiplexing matrix convolution requiring fast weight modulation. Additionally, a previously demonstrated 60-channel DFB laser array with high wavelength precision[49] can substantially increase the computation speed due to its high parallelism. In such a configuration, a computation speed of 2×60×10 G = 1.2 TOPS can be expected. In addition, combined with the time-multiplexed spike encoding proposed in ref.[28], it has great potential to realize large-scale photonic spiking neurons array that is needed for a real-world problem.

Compared to silicon photonics-based weighting elements, the proposed DFB-SA laser offers several advantages. Firstly, it eliminates the loss issue as gain amplification can be easily achieved in the InP integration platform. This characteristic is particularly beneficial for the implementation of multi-layer photonic SNNs. On the other hand, weight control using thermo-optic phase shifters in the silicon photonics platform is limited to a tuning rate in the kHz range. In contrast to the weighting devices based on the VCSEL biased below the lasing threshold[53,54], the DFB-SA laser is biased above the lasing threshold but below the self-pulsating threshold. This allows for efficient spike amplification using the gain provided by the DFB-SA laser, which is crucial for cascaded propagation of weighted spikes. Moreover, the DFB-SA laser enables the simultaneous realization of both linear weighting and activation functions in a single device. This capability avoids frequent OE/EO and AD/DA conversions among different layers in a photonic weighting and electronic activation architecture, and alleviates the inherent challenge of optical coupling that arises when attempting to implement large-scale all-optical SNNs through hybrid integration of weighting and activation elements fabricated using different materials and devices.

In conclusion, we have proposed and successfully fabricated a novel photonic spiking neuron chip based on an integrated DFB-SA laser, which enables simultaneous spike activation and linear weighting functions for the first time. This chip has a simple structure and can be readily integrated on a large scale using commercially mature fabrication processes available in photonics foundries. By adjusting the gain current applied to the DFB-SA laser, the chip can be flexibly reconfigured to function as a linear weighting device or nonlinear spike activation device. The fully-functional neuromorphic prototypical system, comprising DFB-SA lasers and FP-SA laser, successfully performed spike-based AND as well as OR operations. Furthermore, using a fabricated four-channel DFB-SA laser array, we benchmarked the hardware-software collaborative inference on the MNIST dataset, achieving an inference accuracy of 87%. Overall, our work demonstrates the potential of the integrated DFB-SA laser for advancing the field of photonic neuromorphic computing, offering scalability of fully-functional integration of lossless multilayer or deep photonic SNN in a single integrated chip.

References

1. Indiveri G, Liu SC. Memory and information processing in neuromorphic systems. *Proc IEEE* **103**, 1379–1397 (2015).
2. Roy K, Jaiswal A, Panda P. Towards spike-based machine intelligence with neuromorphic computing. *Nature* **575**, 607–617 (2019).
3. Marković D, Mizrahi A, Querlioz D, Grollier J. Physics for neuromorphic computing. *Nat Rev Phys* **2**, 499–510 (2020).
4. Nawrocki RA, Voyles RM, Shaheen SE. A mini review of neuromorphic architectures and implementations. *IEEE Trans Electron Devices* **63**, 3819–3829 (2016).
5. Schuman CD, Potok TE, Patton RM, Birdwell JD, Dean ME et al. A survey of neuromorphic computing and neural networks in hardware. arXiv preprint arXiv: 1705.06963 (2017).
6. Painkras E, Plana LA, Garside J, Temple S, Galluppi F et al. SpiNNaker: a 1-W 18-core system-on-chip for massively-parallel neural network simulation. *IEEE J Solid-State Circuits* **48**, 1943–1953 (2013).
7. Benjamin BV, Gao PR, McQuinn E, Choudhary S, Chandrasekaran AR et al. Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proc IEEE* **102**, 699–716 (2014).
8. Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).
9. Shen JC, Ma D, Gu ZH, Zhang M, Zhu XL et al. Darwin: a neuromorphic hardware co-processor based on spiking neural networks. *Sci China Inform Sci* **59**, 1–5 (2016).
10. Davies M, Srinivasa N, Lin TH, Chinya G, Cao YQ et al. Loihi: a neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
11. Pei J, Deng L, Song S, Zhao MG, Zhang YH et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **572**, 106–111 (2019).
12. Wetzstein G, Ozcan A, Gigan S, Fan SH, Englund D et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
13. Shastri BJ, Tait AN, Ferreira de Lima T, Pernice WHP, Bhaskaran H et al. Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics* **15**, 102–114 (2021).
14. Zhou HL, Dong JJ, Cheng JW, Dong WC, Huang CR et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci Appl* **11**, 30 (2022).
15. Huang CR, Sorger VJ, Miscuglio M, Al-Qadasi M, Mukherjee A et al. Prospects and applications of photonic neural networks. *Adv Phys X* **7**, 1981155 (2022).
16. Qi HX, Du ZC, Hu XY, Yang JY, Chu SS et al. High performance integrated photonic circuit based on inverse design method. *Opto-Electron Adv* **5**, 210061 (2022).
17. Li CH, Du W, Huang YX, Zou JH, Luo LZ et al. Photonic synapses with ultralow energy consumption for artificial visual perception and brain storage. *Opto-Electron Adv* **5**, 210069 (2022).
18. Jiao SM, Liu JW, Zhang LW, Yu FH, Zuo GM et al. All-optical logic gate computing for high-speed parallel information processing. *Opto-Electron Sci* **1**, 220010 (2022).
19. Meng XY, Zhang GJ, Shi NN, Li GY, Azaña J et al. Compact optical convolution processing unit based on multimode interference. *Nat Commun* **14**, 3000 (2023).
20. Zhang F, Guo YH, Pu MB, Chen LW, Xu MF et al. Meta-optics empowered vector visual cryptography for high security and rapid decryption. *Nat Commun* **14**, 1946 (2023).
21. He C, Zhao D, Fan F, Zhou HQ, Li X et al. Pluggable multitask diffractive neural networks based on cascaded metasurfaces. *Opto-Electron Adv* **7**, 230005 (2024).
22. Maass W. Networks of spiking neurons: the third generation of neural network models. *Neural Netw* **10**, 1659–1671 (1997).
23. Gütig R, Sompolinsky H. The tempotron: a neuron that learns spike timing–based decisions. *Nat Neurosci* **9**, 420–428 (2006).
24. Ponulak F, Kasiński A. Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting. *Neural Comput* **22**, 467–510 (2010).
25. Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WHP. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **569**, 208–214 (2019).
26. Xiang SY, Ren ZX, Song ZW, Zhang YH, Guo XX et al. Computing primitive of fully VCSEL-based all-optical spiking neural network for supervised learning and pattern classification. *IEEE Trans Neural Netw Learn Syst* **32**, 2494–2505 (2021).
27. Jha A, Huang CR, Peng HT, Shastri B, Prucnal PR. Photonic spiking neural networks and graphene-on-silicon spiking neurons. *J Lightwave Technol* **40**, 2901–2914 (2022).
28. Xiang SY, Shi YC, Guo XX, Zhang YH, Wang HJ et al. Hardware-algorithm collaborative computing with photonic spiking neuron chip based on an integrated Fabry–Perot laser with a saturable absorber. *Optica* **10**, 162–171 (2023).
29. Tait AN, Ferreira de Lima T, Zhou E, Wu AX, Nahmias MA et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci Rep* **7**, 7430 (2017).
30. Shen YC, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* **11**, 441–446 (2017).
31. Cheng ZG, Ríos C, Pernice WHP, Wright CD, Bhaskaran H. On-chip photonic synapse. *Sci Adv* **3**, e1700160 (2017).
32. Zhou HL, Zhao YH, Wang X, Gao DS, Dong JJ et al. Self-configuring and reconfigurable silicon photonic signal processor. *ACS Photonics* **7**, 792–799 (2020).
33. Feldmann J, Youngblood N, Karpov M, Gehring H, Li X et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
34. Xu XY, Tan MX, Corcoran B, Wu JY, Boes A et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
35. Xu SF, Wang J, Shu HW, Zhang ZK, Yi SC et al. Optical coherent dot-product chip for sophisticated deep learning regression. *Light Sci Appl* **10**, 221 (2021).
36. Zhang H, Gu M, Jiang XD, Thompson J, Cai H et al. An optical neural chip for implementing complex-valued neural network. *Nat Commun* **12**, 457 (2021).
37. Guo XH, Xiang JL, Zhang YJ, Su YK. Integrated neuromorphic photonics: synapses, neurons, and neural networks. *Adv Photonics Res* **2**, 2000212 (2021).
38. Cheng JW, Zhao YH, Zhang WK, Zhou HL, Huang DM et al. A small microring array that performs large complex-valued matrix-vector multiplication. *Front Optoelectron* **15**, 15 (2022).
39. Prucnal PR, Shastri BJ, Ferreira de Lima T, Nahmias MA, Tait AN. Recent progress in semiconductor excitable lasers for photonic spike processing. *Adv Opt Photonics* **8**, 228–299 (2016).

40. Robertson J, Wade E, Kopp Y, Bueno J, Hurtado A. Toward neuromorphic photonic networks of ultrafast spiking laser neurons. *IEEE J Sel Top Quantum Electron* **26**, 7700715 (2020).
41. Zhang YH, Robertson J, Xiang SY, Hejda M, Bueno J et al. All-optical neuromorphic binary convolution with a spiking VCSEL neuron for image gradient magnitudes. *Photonics Res* **9**, B201–B209 (2021).
42. Nahmias MA, Shastri BJ, Tait AN, Prucnal PR. A leaky integrate-and-fire laser neuron for ultrafast cognitive computing. *IEEE J Sel Top Quantum Electron* **19**, 1800212 (2013).
43. Shastri BJ, Nahmias MA, Tait AN, Rodriguez AW, Wu B et al. Spike processing with a graphene excitable laser. *Sci Rep* **6**, 19126 (2016).
44. Chakraborty I, Saha G, Sengupta G, Roy K. Toward fast neural computing using all-photonic phase change spiking neurons. *Sci Rep* **8**, 12980 (2018).
45. Selmi F, Braive R, Beaudoin G, Sagnes I, Kuszelewicz R et al. Relative refractory period in an excitable semiconductor laser. *Phys Rev Lett* **112**, 183902 (2014).
46. Peng HT, Angelatos G, Ferreira de Lima T, Nahmias MA, Tait AN et al. Temporal information processing with an integrated laser neuron. *IEEE J Sel Top Quantum Electron* **26**, 5100209 (2020).
47. Xiang JL, Zhang YJ, Zhao YT, Guo XH, Su YK. All-optical silicon microring spiking neuron. *Photonics Res* **10**, 939–946 (2022).
48. Zheng DZ, Xiang SY, Guo XX, Zhang YH, Gu BL et al. Experimental demonstration of coherent photonic neural computing based on a Fabry–Perot laser with a saturable absorber. *Photonics Res* **11**, 65–71 (2023).
49. Shi YC, Li SM, Chen XF, Li LY, Li JS et al. High channel count and high precision channel spacing multi-wavelength laser array for future PICs. *Sci Rep* **4**, 7377 (2014).
50. Shi YC, Xiang SY, Guo XX, Zhang YH, Wang HJ et al. Photonic integrated spiking neuron chip based on a self-pulsating DFB laser with a saturable absorber. *Photonics Res* **11**, 1382–1389 (2023).
51. Beck H, Yaari Y. Plasticity of intrinsic neuronal properties in CNS disorders. *Nat Rev Neurosci* **9**, 357–369 (2008).
52. The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/.
53. Alanis JA, Robertson J, Hejda M, Hurtado A. Weight adjustable photonic synapse by nonlinear gain in a vertical cavity semiconductor optical amplifier. *Appl Phys Lett* **119**, 201104 (2021).
54. Robertson J, Alanis JA, Hejda M, Hurtado A. Photonic synaptic system for MAC operations by interconnected vertical cavity surface emitting lasers. *Opt Mater. Express* **12**, 1417–1426 (2022).

Acknowledgements

We are grateful for financial supports from National Key Research and Development Program of China (2021YFB2801900, 2021YFB2801901, 2021YFB2801902, 2021YFB2801904); National Natural Science Foundation of China (No. 61974177); National Outstanding Youth Science Fund Project of National Natural Science Foundation of China (62022062); The Fundamental Research Funds for the Central Universities (QTZX23041). We also would like to thank Prof. Xun Li for his helpful discussions.

Author contributions

S. Y. Xiang proposed the idea, Y. C. Shi, X. F. Chen, and H. L. Wang fabricated the samples. Y. H. Zhang, X. X. Guo, L. Zheng performed system experiments. Y. N. Zhang, D. Z. Zheng tested the chips. Y. N. Han, Z. W. Song, T. Zhang developed the algorithms. S. Y. Xiang, Y. C. Shi, X. J. Zhu, M. Qiu, and Y.C. Shen analyzed the data. S. Y. Xiang, Y. C. Shi prepared the manuscript. S. Y. Xiang, W. H. Zheng, and Y. Hao supervised the overall projects. All the authors analyzed and discussed the results.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at https://doi.org/10.29026/oea.2023.230140