Supplemental Document

OPTICA

Hardware-algorithm collaborative computing with photonic spiking neuron chip based on an integrated Fabry–Perot laser with a saturable absorber: supplement

SHUIYING XIANG,^{1,*,†} ^(b) YUECHUN SHI,^{2,3,7,†} XINGXING GUO,¹ YAHUI ZHANG,¹ HONGJI WANG,³ DIANZHUANG ZHENG,¹ ZIWEI SONG,¹ YANAN HAN,¹ SHUANG GAO,¹ SHIHAO ZHAO,¹ BILING GU,¹ HAILING WANG,⁴ XIAOJUN ZHU,⁵ ^(b) LIANPING HOU,⁶ ^(b) XIANGFEI CHEN,³ WANHUA ZHENG,⁴ ^(b) XIAOHUA MA,¹ AND YUE HAO¹

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, Ningbo 315202, China
 ³ Key Laboratory of Intelligent Optical Sensing and Manipulation, Ministry of Education, the National Laboratory of Solid State Microstructures, the College of Engineering and Applied Sciences, Institute of Optical Communication Engineering, Nanjing University, Nanjing 210023, 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, Jiangsu 226019, China
 ⁶ James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK
 ⁷ e-mail: yuechun-shi@ylab.ac.cn
 ^{*} Corresponding author: syxiang@xidian.edu.cn
 [†] These authors contributed equally to this work.

This supplement published with Optica Publishing Group on 24 January 2023 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.21711182

Parent Article DOI: https://doi.org/10.1364/OPTICA.468347

Supplementary Material

Hardware-algorithm collaborative computing with photonic spiking neuron chip based on integrated Fabry– Pérot laser with saturable absorber

Shuiying Xiang^{1, 2*}, Yuechun Shi^{3, 4*}, Xingxing Guo¹, Yahui Zhang^{1, 2}, Hongji

Wang⁴, Dianzhuang Zheng¹, Ziwei Song¹, Yanan Han¹, Shuang Gao¹, Shihao

Zhao¹, Biling Gu¹, Hailing Wang⁵, Xiaojun Zhu⁶, Lianping Hou⁷, Xiangfei Chen

⁴, Wanhua Zheng⁵, Xiaohua Ma² & Yue Hao²

¹ State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

² State Key Discipline Laboratory of Wide Bandgap Semiconductor Technology, School of Microelectronics, Xidian University, Xi'an 710071, China

³Yongjiang laboratory, No. 1792 Cihai South Road Ningbo, 315202, China.

⁴Key Laboratory of Intelligent Optical Sensing and Manipulation, Ministry of Education, the National Laboratory of Solid State Microstructures, the College of Engineering and Applied Sciences, Institute of Optical

Communication Engineering, Nanjing University, Nanjing 210023, China

⁵Laboratory of Solid-State Optoelectronics Information Technology, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China

⁶School of Information Science and Technology, Nantong University, Nantong, Jiangsu, 226019, China

⁷ James Watt School of Engineering, University of Glasgow, Glasgow, G12 8QQ, UK

*Corresponding author: <u>syxiang@xidian.edu.cn</u>; <u>yuechun-shi@ylab.ac.cn</u>;

1. Details of the temporal spike encoding.

The spike encoding for each input pattern can be expressed as $t_e(m)=I(m)\times[x+y+5]ns$, in which x and y are the subscript index of elements in pixel matrix, respectively. I(m) represents the *m*-th column pixel intensity encoded by the PRE-m neuron, and the value is 0 or 1 corresponding to the pixel of white or black. For example, for input pattern "2", pixel intensity of each columns are I(I)=[1,0,1,1,1], I(2)=[1,0,1,0,1], I(3)=[1,0,1,0,1] and I(4)=[1,1,1,0,1], respectively. And the corresponding encoded spiking timings are $t_e(I)=[7, 9, 10, 11]$ ns, $t_e(2)=[8, 10, 12]$ ns, $t_e(3)=[9, 11, 13]$ ns, $t_e(4)=[10, 11, 12, 14]$ ns.

2. Details of the model for the photonic spiking neuron.

In the simulation model, the spontaneous emission coupling factor $\beta = 1 \times 10^{-4}$, output power coupling coefficient $\eta_c = 0.4$, bimolecular recombination term $B_r = 10 \times 10^{-6} \text{ m}^3 \text{s}^{-1}$, photon lifetime $\tau_{ph} = 4.8 \text{ps}$. The other parameter definitions and values for the model are presented in Table S1. The subscripts *a* and *s* stand for the gain and absorber regions, respectively. With these parameters, the rate equations are numerically solved by using the fourth-order Runge-Kutta method.

Parameter	Gain region	Absorber region
Cavity volume	$V_a = 2.4 \times 10^{-18} \mathrm{r}$	$V_s = 2.4 \times 10^{-18} \mathrm{m}^3$
Confinement factor	$\Gamma_a = 0.06$	$\Gamma_s = 0.05$
Carrier lifetime	$\tau_a = 1$ ns	$\tau_s = 100 \mathrm{ps}$
Differential gain/loss	$g_a = 2.9 \times 10^{-12} \mathrm{m}$	$g_s = 14.5 \times 10^{-12} \mathrm{n}$
Transparency carrier	$n = 1.1 \times 10^{24} \text{m}$	$n = 0.89 \times 10^{24} \mathrm{m}^3$
density	$n_{0a} = 1.1 \times 10^{-11}$	$n_{0s} = 0.89 \times 10^{-111}$
Bias current	$I_a = 2 \text{mA}$	$I_s = 0 \text{mA}$

Table S1. The parameters used in the simulation [1-2].

3. Refractory period of FP-SA under different conditions.

We performed extensive experimental measurements of the refractory period. Ten input pulse pairs with different interspike interval (ISI) were designed, and a single input pulse is added as a reference pulse to ensure that each single stimulus pulse exceeds the excitable threshold. Here, the ISI includes 0.1ns, 0.2ns, 0.3ns, 0.4ns, 0.5ns, 0.6ns, 0.7ns, 0.8ns, 0.9ns and 1ns, as shown in Fig. S1 (a1). The response shown in (a2) indicates that, the first 5 pulse pairs each only triggers a single response spike. For the rest cases of ISI, two spikes could be triggered for each pulse pair. That is to say, when ISI is relatively small, the FP-SA cannot emit a spike in a short time after it just responded a spike to the preceding stimulus pulse, as the carrier in the gain region is not fully recovered. But when the ISI is sufficiently large, due to the recovery process of carrier, the gain section takes enough time to fully recover its gain, and the second stimulus pulse can also trigger a spike. Thus, the refractory period in Fig.S1 (a2) is between 0.5ns and 0.6ns. For simplicity, we consider here the refractory period is 0.5ns.

In fact, the refractory period is related with the carrier recovery time. Under different operation conditions, such as gain current, reverse voltage and injection power, which all can affect the carrier number in the gain region, the carrier recovery time is different, leading to different refractory period. We further presented in Figs. S1 (b1-b2) and (c1-c2) under different operation conditions. It can be seen that the refractory period is slightly different, which is about 0.4ns in (b2) and is 0.3ns in (c2).



Fig.S1. The refractory period of the FP-SA. (a1, b1, c1) correspond to input stimuli and (a2, b2, c2) correspond to response of the FP-SA. (a1-a2) with P_{inj} =63.27 µW, V_{SA} =-2.0092V. (b1-b2) P_{inj} =63.9 µW, V_{SA} =-0.66V (c1-c2) P_{inj} =68.4 µW, V_{SA} =0V. The rest parameters are the same for three cases, with I_G =52.0mA, λ_{FP-SA_peak} =1555.510nm, λ_{inj} =1562.874nm.

4. Inhibitory dynamics of FP-SA.

To demonstrate the inhibitory dynamics [3-5], we properly adjusted the operation parameters to make the FP-SA operates at periodic self-pulsation regime, and then injected modulated external stimulus signal to inhibit the spike generation. The result is shown in Fig.S2. It can be seen that, the period pulse train can be inhibited under the high intensity stimulus, due to the injection-locking effect [5]. As presented in (c), the temporal maps plotting the response of laser neuron to the arrival of 100 consecutive external stimuli indicates that the inhibition of spikes is repeatable. The optical spectra is further displayed in (d). Thus, the inhibitory dynamics can also be observed in the fabricated FP-SA.



Fig.S2. The inhibitory dynamics of FP-SA, (a) represents the stimulus signal, (b) denotes the response of the FP-SA neuron, (c) denotes temporal maps plotting the response of laser neuron to the arrival of 100 consecutive external stimuli, (d) represents the optical spectra. With I_G =52.0mA, V_{SA} =-2.01V, P_{inj} =197.46 μ W, λ_{ini} =1560.345nm.

5. The effect of injection power and wavelength detuning on the neuron-like dynamics.

In experiments, it is found that the injection power range that leads to successful neuron-like dynamics is varied with the injection wavelength. We performed extensive measurements on the injected power range for different wavelength separation.

It was found that the FP-SA could operate as a photonic spiking neuron when the injection wavelength was set close to or slightly larger than one of the longitudinal mode wavelength λ_m that is away from the peak wavelength $\lambda_{FP-SA-peak}$.

Through our experimental measurements, the separation between λ_m and $\lambda_{FP-SA-peak}$ that is greater than 2nm is suggested.

At first, we considered a fixed λ_m that is away from the peak wavelength, and varied the injection wavelength of the TL. Here, the wavelength detuning $\Delta \lambda = \lambda_{inj} - \lambda_m$ was varied from 0.01nm to 0.03nm. The measured values of minimum and maximum injection power that lead to spiking threshold response are presented in Table S2. It can be seen that, the injection power range that leads to controllable spike threshold response varied with the wavelength detuning. In general, the

minimum injection power required to realize the threshold response is increased with the increase of wavelength detuning. Besides, we also consider the case when the external stimulus light is injected at different longitudinal modes. For each longitudinal mode wavelength λ_m , we keep the wavelength detuning as a constant $\Delta \lambda = \lambda_{inj} - \lambda_m = 0.02$ nm. The measured values of minimum and maximum injection power that lead to spiking threshold response are presented in Table S3. It can be seen that, when the separation between the injection wavelength and the peak wavelength becomes large, the required minimum injection power is also increased.

Thus, in experiments, there is a range of injection power that makes FP-SA operate as a photonic spiking neuron, and this range is different for different wavelength detuning conditions. Note, for different gain current and reverse voltage conditions, the injection power range will also be slightly different, but the overall trend is not changed.

Common conditions: $I_G = 55.3 \text{ mA}$, $V_{SA} = -2.689 \text{ V}$, $\lambda_{FP-SA_peak} = 1558.360 \text{ nm}$, $\lambda_m = 1561.432 \text{ nm}$						
λ_{inj}	1561.442nm	1561.447nm	1561.452nm	1561.457nm	1561.462nm	
$\Delta \lambda = \lambda_{_{inj}} - \lambda_{_{m}}$	0.01 nm	0.015 nm	0.02 nm	0.025 nm	0.03 nm	
P_{inj_\min}	53.253 µW	57.267 µW	76.293 µW	$77.427\mu\mathrm{W}$	185.04 µW	
P _{inj_max}	72.486 µW	74.025 μW	107.46 µW	108.54 μW	217.26 µW	

Fable S2.	. The minimum	and maximu	n injection	power that	lead to	spike	threshold	response	for diff	erent	cases (of
wavelength detuning.												

r

 Table S3. The minimum and maximum injection power that lead to spike threshold response when injecting near different longitudinal modes.

Common conditions: $I_G = 55.3 \text{ mA}$, $V_{SA} = -2.689 \text{ V}$, $\lambda_{FP-SA_peak} = 1558.360 \text{ nm}$, $\Delta \lambda = \lambda_{inj} - \lambda_m = 0.02 \text{ nm}$						
λ_m	1561.192nm	1561.432nm	1561.672nm	1561.712nm	1562.152nm	
λ_{inj}	1561.212nm	1561.452nm	1561.692nm	1561.732nm	1562.172nm	
P _{inj_min}	54.675 µW	82.305 µW	136.53 µW	176.13 µW	191.43 µW	
P _{inj_max}	81.648 µW	118.53 µW	171 µW	218.43 µW	228.69 µW	

6. Simulation results of pattern recognition with PSNN for patterns "1234".

During the training process, an epoch means all the samples are feed to the network during the training process. When each neuron fires a spike or does not fire as defined by the target for all the input patterns, and the firing state maintains for at least 100 consecutive epochs, it is regarded as training convergence. The convergence epoch is defined as the first epoch of the 100 consecutive epochs. Here, the accuracy reached 100% during the training process.

As shown in Fig. S3, each row is the simulation results of one number pattern. As shown in the first row, (a1) is the representation of number "1" in the 5×4 pixel matrix, the corresponding spatial-temporal spike encoding results of four pre-synaptic neurons are shown in (b1). (c1) and (d1) display the weighted signals that are injected into each POST and the corresponding responses of each POST for pattern "1". (a2-d2), (a3-d3) and (a4-d4) are the simulation results of the number pattern "2", "3" and "4", respectively.



Fig. S3. Simulation results of pattern recognition with PSNN for patterns "1" "2" "3" "4". The representation of number patterns "1" (a1), "2" (a2), "3" (a3), and "4" (a4) in the 5×4 pixel matrix, respectively. (b1, b2, b3, b4) The spatial-temporal spike encoding results of four PREs corresponding to (a1, a2, a3, a4). (c1, c2, c3, c4) represent the inputs of each POST for the pattern "1", "2", "3", and "4", respectively. (d1, d2, d3, d4) the response of each POST corresponding to (c1, c2, c3, c4).

7. Reproducible experimental results of pattern recognition with PSNN for patterns "1234".

Note, in experiments, the electronic noise is inevitable in the AWG, PD, and OSC, as well as the environment variation and the FP-SA laser neuron's noise. To demonstrate the robustness to noise of the pattern recognition results, we further presented the experimental colour-coded temporal maps plotting superimposed time series of the responses corresponding to 500 consecutive arriving stimuli events for the four patterns in Fig. S4. It can be seen that the same spiking response is obtained for each pattern for the 500 consecutive stimuli events. Hence, reproducible pattern classification results can be achieved with the fabricated photonic spiking neuron based on the FP-SA.



Fig. S4. Temporal maps plotting the response of photonic spiking laser neuron to the arrival of 500 consecutive external stimuli. (a) - (d) corresponds to pattern "1", "2", "3", and "4", respectively.

8. The measured optical spectra for both FP-SA1 and FP-SA2.

To construct a cascaded photonic SNN, we first let each FP-SA work alone, and inject the same external optical stimulus to FP-SA1 and FP-SA2. By adjusting the gain current and reverse voltage, both of FP-SA1 and FP-SA2 can exhibit neuron-like spiking response to the same stimulus. Then the output of FP-SA1 is injected to FP-SA2, and for the injection power between 127.8µWand 184.5µW, the cascadability can be achieved. The optical spectra for both FP-SAs when the cascadability is achieved is presented in Fig. S5.



Fig.S5.The optical spectrum of (a) FP-SA1 and (b) FP-SA2 in the cascaded configuration.

9. Simulation results of pattern recognition with PSNN for 'XDU' and 'NJU' tasks.

Here, the network consists of 5 PREs and 3 POSTs for the pattern recognition task of "XDU". For the pattern recognition task of "NJU", the network with 10 PREs and 3 POSTs is employed.

Figures S6 (a) and (c) are the training processes of patterns "XDU" and "NJU", respectively. We can see that the training convergences for "XDU" and "NJU" are achieved at 26 epochs and 16 epochs, respectively. Figures S6 (b) and (d) show the weights after training convergence. As shown in Figs. S7 (a1-a4) and S8 (a1-a4), the input patterns are represented by a 5×5 pixel matrix. Each row shows the simulation results of one pattern, including the spatial-temporal spike encoding results of PREs, the weighted signals that are injected into each POST and the corresponding responses of each POST.



Fig. S6. (a) The training process and (b) weight after training convergence for pattern "XDU". (a) The training process and (b) weight after training convergence for pattern "NJU".



Fig. S7. The simulation results of pattern recognition for pattern "XDU".



Fig. S8. The simulation results of pattern recognition for pattern "NJU".

10. Discussion on the energy consumption, operation speed and potential scalability.

Energy consumption. In artificial neural networks, multiply-and-accumulate (MAC) operations are adopted for benchmarking. However, there is no standard benchmarking method for SNN consumption as SNN computations are based on spike events. Similar to Ref. [6], we consider an operation as one spike event in the SNN. One spike event means a neuron receives all the stimuli spikes from the previous layer and then generates a response spike.

In our experiment, the FP-SA neuron is subject to modulated optical pulse injection. We take an example as follows to estimate the energy consumption. The power of optical injection is 83.835μ W with all "0" inputs. The power of optical injection is 85.059μ W with pulse inputs. Here, pulses occupy 1.67% of the time in a certain time slot. That is to say, the average power provided by the pulse is 1.224μ W (85.059μ W- 83.835μ W= 1.224μ W). The peak dynamics power of pulse can be calculated as 73. 293μ W (1.224μ W/1.67%=73. 293μ W). The width of pulse of 0.1ns, and thus, the pulse provides the energy of 7.329fJ (73.293μ W ×0.1ns=7.329fJ). In the experiment, a spike can be triggered in FP-SA neuron by a single input pulse mentioned above. Thus, the energy efficiency of the FP-SA neuron is 7.329fJ/spike. As for the static power in our neurons, the power consumed is caused by optical injection with 83.835μ W.

We further calculate the energy efficiency in terms of energy per MAC operation. In our experiment, time-multiplexed spike temporal encoding mechanism is used, and a MZM modulator is employed to realize the weighting function. The drive voltage and currents of the MZM modulator is 1.6V and 1mA. The electrical power consumed by a single modulator is 1.6W. Thus, the electrical power consumed in our experiment is $P_{total} = P_s / \eta_{wp} + P_{MZM}$ [7], where P_s stands for the optical power at the input of an axon, η_{wp} denotes wall-plug efficiency of a source laser. P_{MZM} is the electrical power consumed by the MZM modulator. Here, P_s is 10mW, η_{wp} is considered typically as 20%. P_{MZM} is 1.6mW. Thus, $P_{total} = 51.6$ mW. Here, the modulation speed of the MZM modulator is 10GHz, Thus, we think the throughput is T=10GMAC/s. Thus, energy efficiency can be calculated as $E_{MAC} = P_{total} / T = 5.16 \text{ pJ} / MAC$ [7]. Note, in electronic

approaches, the methods of benchmarking based on calculating the conventional multiply-accumulate (MAC) operation show 1300pJ/MAC in TrueNorth and 226.pJ/MAC in Loihi [6]. Thus, the proposed scheme exhibits better performance in terms of energy efficiency.

Operation speed. In our experiment, the maximum rate of external stimuli signal generated by the AWG is 10Gbps. On the one hand, when the input power is relatively low (i.e., single weak pulse cannot trigger a spike), the high speed input pulse can be temporally integrated by the FP-SA, and once the accumulated energy exceeds the spike threshold, a spike can be triggered. In this case, the spike processing speed can be estimated as 10Gbps. On the other hand, when the input power is sufficiently large (i.e., single stimulus pulse energy exceeds the spike threshold), a spike can be trigged upon the FP-SA receives the first stimulus pulse. However, the FP-SA cannot respond spikes successively. Due to the refractory period of the FP-SA (the measured value is about 0.3ns~0.5ns), the FP-SA cannot be triggered another spike within the refractory period. For this case, the FP-SA can process input pulse with rate of 10Gpbs, and its maximum spiking response rate is estimated at about 3.3Gbps (i.e., the inverse of refractory period). Note, the refractory period is related to the carrier recovery time, which will be different under distinct operation conditions.

Potential scalability. On the one hand, the FP-SA can be easily integrated into large-scale bar array. For our fabricated chips, each bar contains about 83 chips. Thus, it has great potential to realize large-scale photonic spiking neurons array, especially with the proposed time-multiplexed temporal spike encoding approach. The scalability is mainly limited by the available packaging technique. On the other hand, it has suggested that the tensor-train decomposed synaptic interconnections could realize large-scale photonic neural networks with reduced hardware resources [8]. Thanks to the advancement of hybrid integration technique, the FP-SA array can be integrated with the synapse network based on the silicon photonics. Thus, by combining the advantages of tensor-train decomposed synaptic interconnection network and the time-multiplexed temporal spike encoding of the proposed FP-SA-based photonic spiking neuron array, it is expected to solve real-scale problems far beyond the hardware scalability limit.

References.

- S. Xiang, Z. Ren, Z. Song, Y. Zhang, X. Guo, G. Han, and Y. Hao, "Computing primitive of fully-VCSELs-based all-optical spiking neural network for supervised learning and pattern classification," IEEE Trans. Neural Netw.Learn. Syst. 32, 2494– 2505 (2021).
- Z. Song, S. Y. Xiang, S. Zhao, Y. Zhang, X. Guo, Y. Tian, Y. Shi, and Y. Hao, "A Hybrid-integrated photonic spiking neural network framework based on an MZI Array and VCSELs-SA," IEEE J. Sel. Top. Quantum Electron., 29, 8300211 (2023).
- J. Xiang, Y. Zhang, Y. Zhao, X. Guo and Y. Su. "All-optical silicon microring spiking neuron," Photon. Res. 10, 939–946 (2022).
- 4. P. Y. Ma, B. J. Shastri, T. F. De Lima, C. Huang, A. N. Tait, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Simultaneous excitatory and inhibitory dynamics in an excitable laser," Opt. Lett. 43, 3802–3805 (2018).
- J. Robertson, T. Ackemann, L. F. Lester and A. Hurtado, "Externally-Triggered Activation and Inhibition of Optical Pulsating Regimes in Quantum-Dot Mode-locked Lasers," Sci. Rep., 8:12515 (2018).
- Y.-J. Lee, M. B. On, X. Xiao, R. Proietti and S. J. Ben Yoo, "Photonic spiking neural networks with event-driven femtojoule optoelectronic neurons based on Izhikevich-inspired model," Opt. Express, 30, 19360-19389 (2022).
- A. R.Totović, G.Dabos, N.Passalis, A.Tefas, and N.Pleros, "Femtojoule per MAC neuromorphic photonics: an energy and technology roadmap," IEEE J. Sel. Top. Quantum Electron., 26, 8800115, 1–15 (2020).
- X. Xiao and S. J. Ben Yoo, "Scalable and compact 3D tensorized photonic neural networks," in Optical Fiber Communication Conference (OFC) 2021, Washington, DC, p. Tu5H.6 (2021).