Reprogrammable All-Optical Ultra-Compact Nonlinear Activation Function for Neuromorphic Computing on Indium Phosphide Membrane on Silicon

Original
Reprogrammable All-Optical Ultra-Compact Nonlinear Activation Function for Neuromorphic Computing on Indium Phosphide Membrane on Silicon / Lechiara, Antonio; Marchisio, Andrea; Jiao, Yuqing; Stabile, Ripalta. - (2025). (2025 IEEE Photonics Society Benelux Symposium Bruxelles (Bel) 20-21 November 2025).

Availability:
This version is available at: 11583/3010967 since: 2026-05-18T10:52:58Z

Publisher:
IEEE Photonic Society

Published
DOI:

Reprogrammable All-Optical Ultra-Compact Nonlinear Activation Function for Neuromorphic Computing on Indium Phosphide Membrane on Silicon

A. Lechiara[1], A. Marchisio[2], Y. Jiao[1], and R. Stabile[1]

[1]Casimir Institute, Eindhoven University of Technology, Eindhoven 5612 AP, The Netherlands

[2]Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy

Email: a.lechiara@tue.nl[1]

Abstract: In this work we present the characterization and numerical study of a reprogrammable all-optical ultra-compact nonlinear activation function for neuromorphic computing, implemented on the Indium Phosphide Membrane on Silicon (IMOS) platform. The device consists of a Mach–Zehnder coupler (MZC) followed by a microring-resonator (MRR)-nested Mach–Zehnder interferometer (MZI). The nonlinear behavior is driven by the MRR, while three independently tunable degrees of freedom - the MZC phase shifter, the MRR resonance, and a post-MRR phase shifter - enable dynamic reprogramming of the transfer function. By exploiting ultra-compact phase shifters, the architecture achieves a footprint of $0.054\,\mathrm{mm}^2$ and a $-5\,\mathrm{dBm}$ triggering power. We experimentally characterize the resonant response and tunability of the device, reporting a thermo-optic tuning efficiency of $0.9598\,\mathrm{nm/mW}$. Without any transfer-function optimization, the device reproduces a Leaky Rectified Linear Unit (ReLU) response, as commonly found in modern artificial neural networks when computational efficiency and sparsity are required. Biasing the phase shifters allows the device to reproduce additional transfer functions such as Radial Basis Functions (RBFs) and sigmoidal shapes. These results highlight the potential of compact, reprogrammable photonic structures on the IMOS platform to realize nonlinear activation functions for neuromorphic computing.

Introduction

Despite the considerable improvement of the efficiency of AI hardware based on electronics, the huge power consumption required for operating the servers, GPUs, cooling systems and accessory equipment still constitutes a major limitation for their scalability and sustainability [1]. In this framework, alternatives to conventional electronics, aiming at mitigating the bottlenecks of these architectures, come into play.

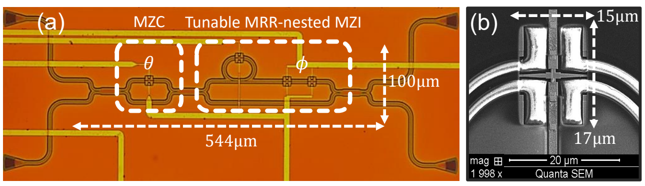


Figure 1: (a) Micrograph of the NLAF. The tunable MZC is followed by a MRR nested on a MZI; (b) Detailed SEM magnification of the ultra-compact PS available on the IMOS platform.

Neuromorphic photonics is a rapidly developing field of research [2] that leverages the high bandwidth and energy efficiency of Photonic Integrated Circuits (PICs) to realize neuromorphic computing systems. In these systems, light—usually a coherent beam generated by a laser source—is used as the communication medium and is manipulated in an analog fashion to implement the desired computation. Central to enabling computation in such architectures is the nonlinear activation function, which allows photonic neural networks to solve complex problems that are not linearly separable and to increase their expressiveness.

[1]A.L. Ph.D. scholarship is funded by the NL-ECO program, through the NWA-ORC programme of the Dutch Research Council (NWO; project number NWA.1389.20.140).

When integrated together with linear optical processors, all-optical nonlinearities can achieve superior speed compared to electronic solutions, making them highly attractive for real-time, high-throughput signal processing. A variety of photonic activation functions has been proposed for optical neural networks, spanning electro-optical implementations [3,4] as well as fully optical nonlinear designs [5–7]. While electro-optical devices typically offer high optical output power and precise tunability, the bandwidth of electrical connections is a limiting factor. On the other hand, all-optical nonlinear functions allow seamless integration with on-chip matrix multiplication units but require more careful tuning.

All-optical nonlinearities are typically weak, meaning that high optical power is necessary to trigger a nonlinear optical response and the output of a neuron is often too weak to drive a subsequent layer of neurons, hence limiting the overall cascadability. In this work, following our preliminary results [8], we present the characterization and numerical study of a reprogrammable, all-optical, ultra-compact nonlinear activation function (NLAF) for neuromorphic computing and low triggering power.

The proposed device, shown in Fig. 1(a), consists of two main sections: a Mach–Zehnder coupler (MZC) and a microring-resonator (MRR)-nested Mach–Zehnder interferometer (MZI) [9]. The effect of the presence of the MRR in the structure is the introduction of a resonant behavior dominated by the characteristics of the ring itself. The nonlinear behavior is driven by the MRR, and the transfer function is reconfigured by biasing the phase shifters (PSs) on the MZC, MRR, and MZI sections.

Methods and Results

The platform of choice to fabricate the NLAF is the InP membrane on silicon (IMOS) platform developed at the Eindhoven University of Technology. The sub-micrometer-thick InP membrane allows devices to have dimensions comparable to SOI while allowing the co-integration of passive components with lasers and optical amplifiers. The IMOS platform is an appealing environment to study the effect of enhanced nonlinearities, with particular interest in third-order ($\chi^{(3)}$) nonlinearities, since the strong optical confinement within the membrane can trigger nonlinearities at an optical power about 5–10 times lower than on silicon [10]. Moreover, we exploit an ultra-compact thermo-optic PS, shown in Fig. 1(b), with a total size of $15 \times 17\,\mu\text{m}^2$, up to one order of magnitude smaller than conventional phase shifters based on metal pads [11].

Both the waveguides and the MRR are designed employing the standard IMOS waveguide cross-section ($400\,\text{nm} \times 300\,\text{nm}$). The MRR has a racetrack geometry with radius $R = 20\,\mu\text{m}$ and coupling length $L_\text{c} = 13.3\,\mu\text{m}$, with a 200 nm gap between them. The size of the device, excluding the grating couplers (GCs), is $0.054\,\text{mm}^2$. To characterize the NLAF, we vertically couple light from a Keysight N7711A tunable laser at an incident angle of 12°, employing a Polarization Controller (PC) to inject light into the PIC. Electrical biasing and the reading of current and voltage of the on-chip phase shifter are performed via a Keithley 2400 power supply. The power injected into the PIC is regulated by an HP 8156A variable optical attenuator (VOA), and the output power is measured with a Santec MPM-210H optical power meter. The coupling losses are estimated to be -6.09 dB per GC, whereas the insertion losses introduced by the VOA amount to -3.3 dB. Further loss contributions along the transmission path are attributed to fiber splitters and mating sleeves. The results of the characterization are reported in Fig. 2.

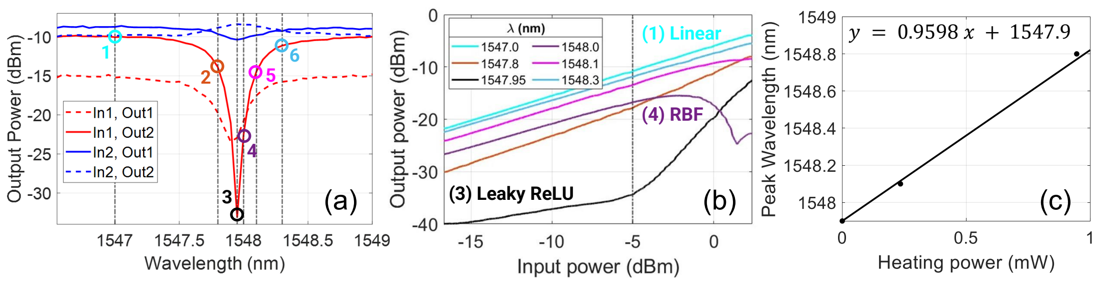
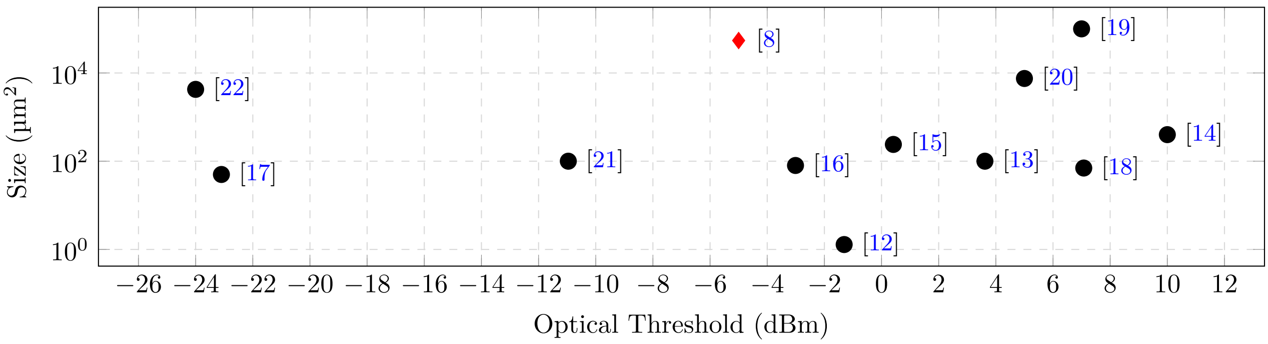


Figure 2: (a) Transmission spectrum of the NLAF; (b) Transfer functions at different values of λ_L; (c) Tunability of the NLAF. The reported input and output power exclude the coupling and VOA losses.

First, we measure the transmission spectrum of the NLAF by sweeping the wavelength of the tunable laser λ_L at a fixed input power $P_\text{in} = 6$ dBm for all four combinations of input and output ports and measuring the output power P_out. As shown in Fig. 2(a), most of the optical power enters the MRR in the upper arm of the MZI when the input 1 and output 2 ports are used. We identify a resonance wavelength $\lambda_\text{res} = 1547.95$ nm associated with an extinction ratio ER ≈ 23.15 dB.

We then proceed to evaluate the passive response of the NLAF by evaluating the P_{out}–P_{in} curves at input 1 and output 2 ports for different λ_L. This approach is equivalent to biasing the PS of the MRR to detune its spectrum. We set the laser source output power to its maximum value 15 dBm and apply a variable attenuation from −20 dB to 0 dB with the VOA. This results in the P_{in} injected at the input GC spanning from −7.29 dBm to 11.77 dBm. In Fig. 2(b), we report the different transfer functions at various wavelengths around λ_{res}, marked in the spectrum of Fig. 2(a) using the same colors.

Far from λ_{res}, i.e., for $\lambda_L = 1547, 1548.3$ nm (cyan and blue curves), very little light is coupled into the MRR and the NLAF response is linear. When approaching resonance, i.e., for $\lambda_L = 1547.8, 1548.1$ nm (orange and magenta curves), part of the optical power interacts within the ring and deviations from linearity appear. Exactly at $\lambda_L = \lambda_{\text{res}}$ (black curve) the transfer function is highly nonlinear and reproduces a leaky ReLU with functional form $f(x) = \max(3x, 0.45x)$, centered around a triggering power ≈ -5 dBm. To the right of the resonance dip and close to it, i.e. for $\lambda_L \gtrsim \lambda_{\text{res}}$ (purple curve), the response reproduces a radial basis function (RBF). Last, in Fig. 2(c), we report the tunability of the device by biasing the MRR's PS to determine the shift of λ_{res} as a function of the heating power, i.e. the product of voltage and current across the PS. From its functional form $y = 0.9598x + 1547.9\text{nm mW}^{-1}$ we retrieve a tunability of $0.9598\,\text{nm mW}^{-1}$. Finally, in the table below we compare our results (red mark) with other approaches. While reprogrammable devices are typically larger in size, our proposal on the IMOS platform ensures a low optical threshold.



Conclusions

We presented our design of an all-optical NLAF for neuromorphic computing, characterized its passive transfer function for different detunings from resonance and measured the tunability. The final comparison table shows the potentiality of our design on the IMOS platform when low optical thresholds and reprogrammability are required. The investigation of the tunability of the NLAF, leveraging the phase shifters of the MZC and MZI sections, and the testing of the performances of a Neural Network employing such transfer functions are foreseen.

References

[1] J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos, "Compute trends across three eras of machine learning," in *2022 international joint conference on neural networks (IJCNN)*. IEEE, 2022, pp. 1–8.

[2] P. R. Prucnal and B. J. Shastri, *Neuromorphic photonics*. CRC press, 2017.

[3] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 4029–4041, 2014.

[4] I. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *IEEE Journal of Selected Topics in Quantum Electronics*, pp. 1–14, 2019.

[5] B. Shi, "Semiconductor optical amplifier-based photonic integrated deep neural networks," Ph.D. dissertation, Eindhoven University of Technology, 2022.

[6] C. Huang, T. Ferreira De Lima, A. Jha, S. Abbaslou, A. N. Tait, B. J. Shastri, and P. R. Prucnal, "Programmable silicon photonic optical thresholder," *IEEE Photonics Technology Letters*, vol. 31, no. 22, pp. 1834–1837, 2019.

[7] G. Mourgias-Alexandris, G. Dabos, N. Passalis, A. Totović, A. Tefas, and N. Pleros, "All-optical WDM recurrent neural networks with gating," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 5, pp. 1–7, 2020.

[8] L. Antonio, A. Marchisio, S. Bin, Y. Wang, A. Carena, J. Yuqing, S. Ripalta *et al.*, "Ultra-compact leaky relu nonlinear function on imos," in *ECOC 2025 Conference Paper*. IEEE, 2025.

[9] A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," *Optics Letters, Vol. 45, Issue 17, pp. 4819-4822*, vol. 45, pp. 4819–4822, 9 2020.

[10] Y. Jiao, J. van der Tol, V. Pogoretskii, J. van Engelen, A. A. Kashi, S. Reniers, Y. Wang, X. Zhao, W. Yao, T. Liu, F. Pagliano, A. Fiore, X. Zhang, Z. Cao, R. R. Kumar, H. K. Tsang, R. van Veldhoven, T. de Vries, E. J. Geluk, J. Bolk, H. Ambrosius, M. Smit, and K. Williams, "Indium phosphide membrane nanophotonic integrated circuits on silicon," *Physica Status Solidi (A) Applications and Materials Science*, vol. 217, pp. 1–12, 2020.

[11] Y. Wang, V. Dolores-Calzadilla, K. A. Williams, M. K. Smit, and Y. Jiao, "Ultra-compact and efficient microheaters on a submicron-thick InP membrane," *Journal of Lightwave Technology*, vol. 41, no. 6, pp. 1790–1800, 2023.

[12] B. Wu, H. Li, W. Tong, J. Dong, and X. Zhang, "Low-threshold all-optical nonlinear activation function based on a ge/si hybrid structure in a microring resonator," *Optical Materials Express*, vol. 12, no. 3, pp. 970–980, 2022.

[13] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019.

[14] K. Liao, C. Li, T. Dai, C. Zhong, H. Lin, X. Hu, and Q. Gong, "Matrix eigenvalue solver based on reconfigurable photonic neural network," *Nanophotonics*, vol. 11, no. 17, pp. 4089–4099, 2022.

[15] Y. Shi, J. Ren, G. Chen, W. Liu, C. Jin, X. Guo, Y. Yu, and X. Zhang, "Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks," *Nature Communications*, vol. 13, no. 1, p. 6048, 2022.

[16] C. Zhong, K. Liao, T. Dai, M. Wei, H. Ma, J. Wu, Z. Zhang, Y. Ye, Y. Luo, Z. Chen *et al.*, "Graphene/silicon heterojunction for reconfigurable phase-relevant activation function in coherent optical neural networks," *Nature Communications*, vol. 14, no. 1, p. 6939, 2023.

[17] C. Chen, Z. Yang, T. Wang, Y. Wang, K. Gao, J. Wu, J. Wang, J. Qiu, and D. Tan, "Ultra-broadband all-optical nonlinear activation function enabled by mote2/optical waveguide integrated devices," *Nature Communications*, vol. 15, no. 1, p. 9047, 2024.

[18] H. Li, B. Wu, W. Tong, J. Dong, and X. Zhang, "All-optical nonlinear activation function based on germanium silicon hybrid asymmetric coupler," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 29, no. 2: Optical Computing, pp. 1–6, 2022.

[19] I. A. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–12, 2019.

[20] A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," *Optics letters*, vol. 45, no. 17, pp. 4819–4822, 2020.

[21] W. Yu, S. Zheng, Z. Zhao, B. Wang, and W. Zhang, "Reconfigurable low-threshold all-optical nonlinear activation functions based on an add-drop silicon microring resonator," *IEEE Photonics Journal*, vol. 14, no. 6, pp. 1–7, 2022.

[22] G. H. Li, R. Sekine, R. Nehra, R. M. Gray, L. Ledezma, Q. Guo, and A. Marandi, "All-optical ultrafast relu function for energy-efficient nanophotonic deep learning," *Nanophotonics*, vol. 12, no. 5, pp. 847–855, 2023. [Online]. Available: https://doi.org/10.1515/nanoph-2022-0137