Programmable tanh- and ReLU-like Optoelectronic Activation Functions for Neuromorphic Photonic Circuits

Christos Pappas[1], Stefanos Kovaios[1], Miltiadis Moralis-Pegios[1], Apostolos Tsakyridis[1], George Giamougiannis[1], Joris Van Kerrebrouck[2], Gertjan Coudyzer[2], Xin Yin[2] and Nikos Pleros[1]

[1]Centre for Interdisciplinary Research and Innovation, Informatics Dept. Aristotle University of Thessaloniki, Greece
[2]IDLab, Department of Information Technology, Ghent University-imec,9052 Ghent, Belgium

Author e-mail address: chripapp@csd.auth.gr

Abstract: We demonstrate reconfigurable tanh- and ReLU-like nonlinear activation functions for incoherent neuromorphic photonics using a balanced photodiode assembled with a programmable electronic TIA chip. Experimental results up to 10 Gb/s line-rates are presented.

1. Introduction

Neuromorphic photonics comes with the promise of low-energy and high-speed Deep Learning applications, having already demonstrated a great range of impressive experimental prototypes [1],[2]. However, these have mainly focused on the deployment of the linear neuron part where Matrix-Vector Multiplications (MVM) are carried out, with the non-linear activation functions still typically realized in the digital electronic domain. Efforts for demonstrating analog all-optical or optoelectronic nonlinear activations [3]-[8] to leverage their high-speed potential and seamlessly cooperate with the analog MVM engine have shown a remarkable potential for realizing a great range of activation functions that are traditionally used in digital Deep Neural Network (DNN) models, including ReLU- [4]-[9], PReLU- [10] and sigmoid-like [3],[4] functions. At the same time, they have also enabled a new portfolio of non-typical nonlinear functions that have already validated their credentials to support high-accuracy DNN models [11]. In all this photonic activation landscape, the tanh nonlinear function that is also among the typical activations utilized in NN architectures has never been demonstrated in the optical or optoelectronic domain. Tanh is following the shape of the sigmoid function but is symmetric around 0, leading in a much faster convergence. However, tanh activations in very deep NNs may suffer from the vanishing gradient problem [12], which can be solved using ReLU-like functions. Having a universal analog device which can be programmed to support tanh as well as alternative activation functions with reconfigurable characteristics can allow for a hardware solution that can meet a great range of NN models and requirements.

In this paper, we demonstrate for the first time, to the best of our knowledge, an analog optoelectronic activation unit that can be programmed to support tanh-, ReLU- and inverted ReLU-like activations up to 10Gb/s line-rates. The non-linear activation integrated circuit (NLA-IC) chip comprises a balanced photodetector (BPD) that is followed by a wire-bonded programmable Transimpedance Amplifier (TIA). The circuit performance has been experimentally evaluated at 2 and 10Gb/s, producing a range of tanh- and ReLU-like activation functions that have an excellent fit to the nonlinear tanh, Exponential Linear Unit (ELU) and inverted ELU mathematical functions and can have their reconfigurable performance characteristics for their slope and operational range.

2. Non-linear Activation Circuit and Experimental setup

The NLA-IC utilized for the experimental demonstration of variant activation functions consists of a silicon photonic BPD wirebonded to a TIA with programmable nonlinear responses, as represented in the block diagram of Fig.1(a). The BPD is fabricated in imec's iSiPP50G silicon photonics platform [13], using two SiGe PIN photodiodes. The

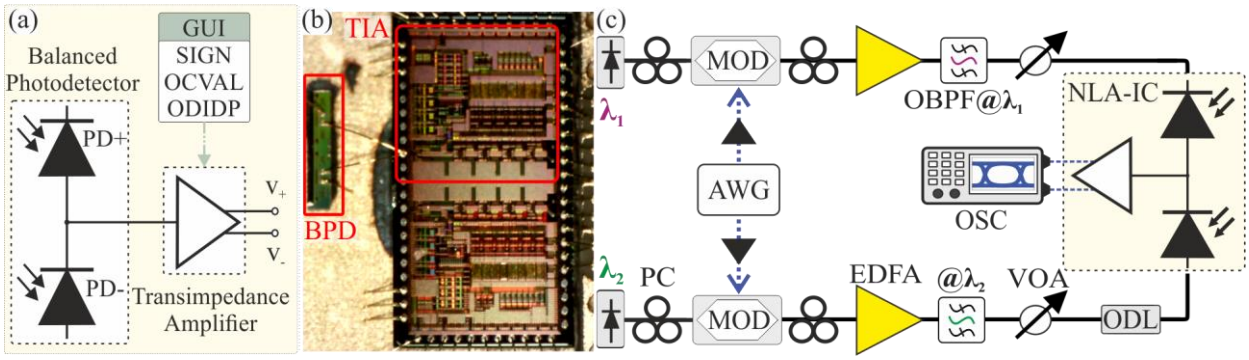

Fig. 1: a) Block diagram of the NLA-IC including a balanced photodetector (BPD) and a transimpedance amplifier (TIA), (b) microscopic image of the chip and (c) experimental setup employed to characterize and obtain the different O-E transfer functions of the NLA-IC.

photodiodes have a nominal responsivity of 0.8A/W and a low dark current ≤40 nA with 1V bias voltage. The TIA chip comprises a single-ended-to-differential (S2D) amplifier, two variable-gain-amplifiers (VGAs) and an output driver (OD) with the total transimpedance gains ranging from 0.5 kΩ to 2.5 kΩ. An on-chip look-up table (LUT) is used to store the digital settings given to the TIA by the GUI, configuring the offset, nonlinearity and gain of the amplifier circuits. Three of the GUI's parameters are in charge of alternating the chip's response, hence the transfer function (TF) of the TIA, named ODIDP, OCVAL and SIGN. The gain and linear range of the output signal is controlled by ODIDP with values in range [0, 7], the offset of the signal by OCVAL in a range [00, 60], which moves the TF onto the x-axis while this movement direction is controlled by the SIGN with available values 0 or 1. The TIA was fabricated in a 0.25 um SiGe:C BiCMOS-technology. Fig. 1(b) shows a microscope photo of the NLA-IC chip that comprises the balanced PD wirebonded to one of the two TIAs employed on the same IC. The dual-TIA IC chip has a total area of 3.62 mm × 1.97 mm [14].

The experimental testbed employed to characterize the NLA-IC as O-E activation functions for neural networks (NNs) is illustrated in Fig. 1(c). Two 10 mW Continuous-Wave (CW) beams encoded on λ_1 =1550 nm and λ_2=1551.6 nm, respectively, were injected into two different $LiNbO_3$ Mach- Zehnder modulators (MOD). The two MODs were driven by an arbitrary waveform generator (AWG) to generate two Non-Return to Zero (NRZ) custom patterns. These two distinctive signals were injected into the NLA chip through the two inputs (positive and negative) of the BPD, with P1 entering the positive input of the BPD and thus carrying positive values and P2 entering the negative input and hence corresponding to negative values. In this way, an electrical single signal that corresponds to P1-P2 is generated at the BPD output. This electrical signal was then fed to the TIA and its two outputs were captured by the oscilloscope (OSC), resulting in a differential signal with a maximum V_{pp} of 400 mV. To maintain proper signal polarization and power levels, polarization controllers (PCs), variable optical attenuators (VOAs) and erbium-doped fiber amplifiers (EDFAs) were used, as shown in the figure. Optical bandpass filters (OBPF) of 1 nm 3dB bandwidth were also utilized for wavelength filtering. An optical delay line (ODL) was used to achieve pattern and bit-level synchronizations between the P1 and P2 optical signals at the BPD inputs.

3. Experimental results

The NLA-IC was experimentally evaluated for different parameter sets of the TIA both at 2Gb/s and 10Gb/s. Fig. 2(a) presents the 2 custom input patterns P1 and P2 that were programmed in software and were sent to the AWG, showing that every signal was selected to consist of 31 different power levels, with P1 and P2 following a gradually increasing and decreasing, respectively, power level along the 31-pulse sequence. This format was selected so that P2 is the time-inverted P1 waveform, so as to allow both positive and negative values to be generated when being

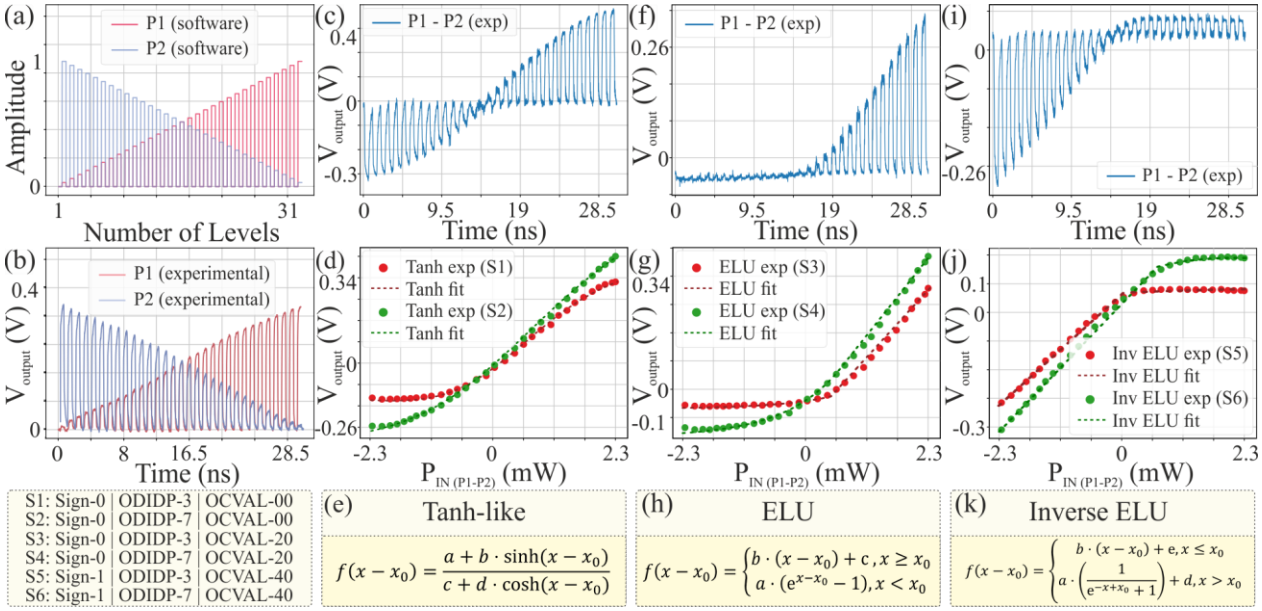

Fig. 2: (a)-(b) Custom P1 and P2 signals at 2 Gb/s, software (a) and experimental (b), consisting of 31 different levels and representing Positive Inputs (P1) and Negative Inputs (P2), (c)-(h) a variety of transfer functions seen from the NLA-IC at 2 Gb/s, (c) P1-P2 experimentally with a set of parameters of the TIA producing the "tanh", (d) two different "tanh" experimental functions along with the theoretical fitting of (e), (f) P1-P2 producing the "ELU", (g) two experimental "ELU" functions fitted with the theoretical function of (h), (i) P1-P2 generate the "inverse ELU" and (j) two experimental "inverse ELU" functions fitted with the theoretical function of (k). S1-S6 represents the different set of parameters of SIGN, ODIDP and OCVAL, for the TIA's response given digitally by software

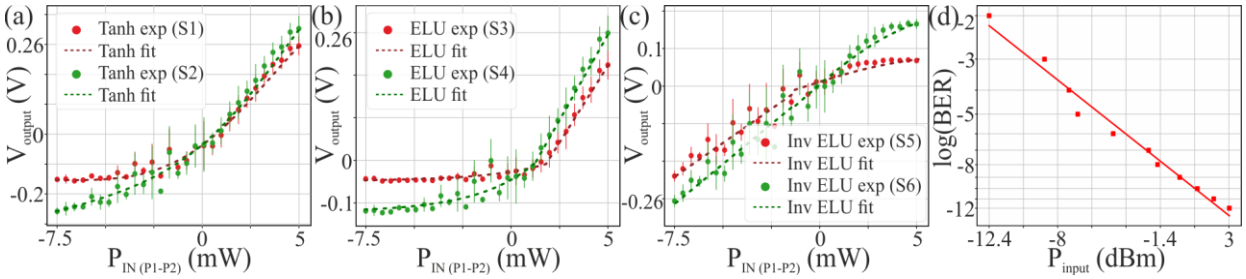

Fig. 3: (a)-(c) Experimental data at 10 Gb/s for "tanh", "ELU" and "inverse ELU" with the respective theoretical fittings of the functions of Fig.2 (e), (h), (k) and (d) BER measurements versus optical input power at 10 Gb/s. S1-S6 represent the same parameter sets as of Fig. 2.

substracted at the BPD outputs. Fig. 2(b) shows the two optical signals generated at the modulator outputs at 2 Gb/s. In Fig. 2(c) the pulse trace of the TIA's output response for the S1 parameters (SIGN: 0, ODIDP: 3 and OCVAL: 00) is illustrated, indicating a tanh-like TF, where the higher negative P2 and positive P1 values get clipped. Fig. 2(d) presents the experimental TFs for S1 and S2 (SIGN: 0, ODIDP: 7 and OCVAL: 00) parameter sets represented with red and green dots, respectively, along with the fitting of the theoretical tanh-like function of Fig. 2(e). The S2 case indicates a change in the amplitude of the output and thus the P1 values do not clip, as in S1 case, since more optical input power was needed. Fig. 2(f) represents the pulse trace of the NLA chip's response for the S3 parameters (SIGN: 0, ODIDP: 3 and OCVAL: 20), which stipulate an ELU TF. In this case only the negative P2 values were clipped at an x_0 value. Fig. 2(g) shows the experimental TFs for S3 and S4 (SIGN: 0, ODIDP: 7 and OCVAL: 20) parameter sets together with the theoretical ELU function fitting of Fig. 2(h). By alternating the gain of the chip, ODIDP parameter, the slope of the ELU and the x_0 point are changing. Finally, in Fig. 2(i) the S5 (SIGN: 1, ODIDP: 3 and OCVAL: 40) set used, resulting in an inverse-ELU chip response where the positive P1 values were clipped at an x_0 value. Fig. 2(j) presents the experimentally obtained TFs for S5 and S6 (SIGN: 1, ODIDP: 7 and OCVAL: 40) sets in addition to the fitting of the theoretical inverse-ELU function, as shown in Fig. 2(k) where once more the higher gain, ODIDP, changes the slope and the x_0 value. In the 2 Gb/s operation, the total power consumption was 187.5 pJ/bit.

Following the evaluation process of 2 Gb/s, Fig 3 shows the experimental activation functions at 10 Gb/s. The input signals P1 and P2 were created in the same way as in the case of 2 Gb/s while the pulse traces indicated a similar behavior as in the lower rate, with a slightly bigger error. The parameter sets remained the same as in case of 2 Gb/s and the experimentally obtained TF curves are illustrated in Figs. 3(a)-(c). Fig. 3(a) shows the experimental TF curves for the tanh-like function with the S1 and S2 parameter sets, along with the fitting of the theoretical tanh-like function of Fig. 2(e). In Fig. 3(b) the experimental TF curves for the ELU chip response are presented, along with the fitting of the theoretical ELU function of Fig. 2(h), while the chip operated with the S3 and S4 parameters. Fig. 3(c) displays the inverse-ELU TFs that were experimentally observed while operating the chip with S5 and S6 parameters along with the theoretical inverse-ELU function fitting of Fig. 2(k). Finally, considering that a 99% NN accuracy can correspond to a bit error rate (BER) of 10^{-2}, the sensitivity of the NLA-IC was evaluated through BER measurements using a PRBS-7 pattern and was found to be -12 dBm at a BER-level of 10^{-2} (Fig. 3(d)). The total power consumption for operation at 10 Gb/s was 37.5 pJ/bit.

Acknowledgements

This work was supported by the European Commission via H2020 Projects PLASMONIAC (871391) and SIPHO-G (101017194).

References

[1] M. Moralis-Pegios et al., "Neuromorphic Silicon Photonics and Hardware-Aware Deep Learning for High-Speed Inference," in JLT, 40, 10, 2022.
[2] B.J. Shastri, et al., "Photonics for artificial intelligence and neuromorphic computing," Nat. Photonics 15, 2021.
[3] G. Mourgias-Alexandris, et al., "An all-optical neuron with sigmoid activation function," Opt. Express 27, 2019.
[4] A. Jha, et al., "Programmable, high-speed all-optical nonlinear activation functions for neuromorphic photonics," OFC, 2021.
[5] I. Williamson, et. al., "Reprogrammable Electro-Optic Nonlinear Activation Functions for Optical Neural Networks," IEEE JSTQE, 26, 2020.
[6] L. Gordon H.Y., et al., "All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning," Nanophotonics, 2022.
[7] B. Wu, et al., "Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator," Opt. Mater. Express 12, 2022.
[8] Y. Zuo, et al., "All-optical neural network with nonlinear activation functions," Optica, 6, 2019.
[9] A. Tait, et. al., "Silicon Photonic Modulator Neuron," Physical Review Applied, 11, 2019.
[10] J. Crnjanski, et. al., "Adaptive sigmoid-like and PReLU activation functions for all-optical perceptron," Optics Letters, 46, 2003, 2021.
[11] N. Passalis, et. al., "Training Deep Photonic Convolutional Neural Networks With Sinusoidal Activations," IEEE TETCI, 5, 2021.
[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, 521, 7533, 2015.
[13] A. Rahim, et al., "Open-access silicon photonics platforms in Europe," IEEE JSTQ, 25, 2019.
[14] G. Coudyzer, et. al., "A 50 Gbit/s PAM-4 Linear Burst-Mode Transimpedance Amplifier," in IEEE PTL, 31, 12, 2019.