



An all-optical neuron with sigmoid activation function

G. MOURGIAS-ALEXANDRIS,^{1,2,*} A. TSAKYRIDIS,^{1,2} N. PASSALIS,¹
A. TEFAS,¹ K. VYRSOKINOS^{2,3} AND N. PLEROS^{1,2}

¹*Department of Informatics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece*

²*Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Center, Buildings A & B, Thessaloniki, 10th km Thessaloniki-Thermi Rd, P.O. Box 8318, GR 57001, Greece*

³*Department of Physics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece*

*mourgias@csd.auth.gr

Abstract: We present an all-optical neuron that utilizes a logistic sigmoid activation function, using a Wavelength-Division Multiplexing (WDM) input & weighting scheme. The activation function is realized by means of a deeply-saturated differentially-biased Semiconductor Optical Amplifier-Mach-Zehnder Interferometer (SOA-MZI) followed by a SOA-Cross-Gain-Modulation (XGM) gate. Its transfer function is both experimentally and theoretically analyzed, showing excellent agreement between theory and experiment and an almost perfect fitting with a logistic sigmoid function. The optical sigmoid transfer function is then exploited in the experimental demonstration of a photonic neuron, demonstrating successful thresholding over a 100psec-long pulse sequence with 4 different weighted-and-summed power levels.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

Deep Learning and Neural Networks (NN) have recently made the turn into highly promising alternative computing platforms, raising expectations to break the barriers in computational efficiencies enforced by Moore's law in conventional computing and lead the race towards beyond von Neumann architectures [1]. Employing millions of electronic-based artificial neurons, spiking neuromorphic processors as the TrueNorth [2] and the SpiNNaker [3] have already been demonstrated showing a tremendous efficiency. Google TPU follows a different approach and employs custom ASICs for matrix multiplications, accelerating the execution of neural network tasks [4] and yielding a peak throughput of 92 TeraOps/Second, improving by 70x the TeraOps/sec/Watt performance compared to state-of-the-art GPUs. In the meantime, the potential of photons has been already identified towards realizing an ultra-fast and power-efficient photonic deep learning processor platform, having the credentials to perform optical matrix multiplications overcoming traditional shortcomings of electronic layouts and reaching operational frequencies up to 100GHz [5]. Linear optics offered by the native interference properties of light have been utilized either in free-space optical circuits [6] or in waveguide-based coherent circuitry [5] for realizing a totally passive optical matrix multiplication unit for neural network applications; at the same time, WDM weighted addition [7] has introduced the entity of wavelength into the field of NN, with micro-ring resonator-based photonic integrated circuits turning into critical elements for weighting functions, enabling also new broadcast-and-weight architectures for supporting both spiking [8] as well as fast Recurrent Neural Network (RNN) layouts [9].

However, the optical neurons that have been demonstrated so far [10–16] have relied on a rather limited set of photonic non-linear processing functions compared to the typical range of activation functions employed in Deep NN (DNN) layouts, despite the variety of photonic non-linear thresholding elements that have been demonstrated as stand-alone modules [17–21]. DNN neurons typically require the ReLU [22], PReLU [23] and variations of the sigmoid transfer function, including tanh and the logistic sigmoid [24] but the vast majority of photonic activation

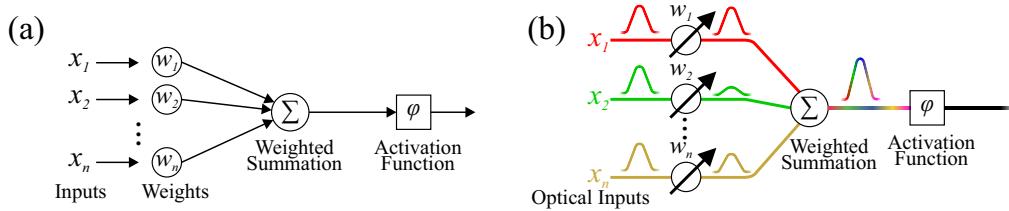


Fig. 1. (a) Artificial neuron, (b) Optical WDM neuron.

functions has been based so far either on close to Heaviside step transfer functions offered usually by directly modulated laser-based neurons [8, 15] or on sinusoidal [9, 21] responses that are usually obtained via an electro-optic modulator driven by the output of a balanced photodiode circuitry [8] or by optical switching elements [13]. Spiking NNs have also been mainly implemented with sinusoidal photonic integrate-and-fire transfer functions typically provided by fiber- or SOA-based thresholding devices [11] or with close to step-wise unitary functions [10]. Although our recent work outlines the training procedure and confirms that optically-enabled sinusoidal activation can lead to converging Deep Convolutional Neural Networks (CNN) [25], sigmoid-based activations continue to comprise the backbone of powerful RNNs [26] and Long-Short Term Memory (LSTM) [24] architectures, being responsible for gating and confining the output between two well-defined levels. Neural network attention mechanisms [27, 28], that provenly improve the accuracy of neural models can be also directly implemented using sigmoid activations [29]. However, there has been still no experimental demonstration reported so far neither for a complete sigmoid optical neuron nor for an optical sigmoid transfer function. The use of threshold-like optical functions that only qualitatively approximate a sigmoid curve has been the mainstream approach in neuromorphic photonic architectures proposed so far [7], but often the conventional sigmoid transfer function offers still a slightly better neural network accuracy [30], while it definitely can't rely on existing training techniques and algorithms, as indicated in our recent work of [25]. This creates also a gap in the photonic deployment perspectives of well-known RNN and LSTM architectures that could employ well-known NN algorithms over photonic hardware platforms. This gap has been already identified in the literature, with recent work reporting on the need for sigmoid optical neurons presenting also simulation-based results for possible photonic implementations [30], but there has been still no experimental evidence of any sigmoid optical neuron presented so far.

In this paper, we present for the first time, to the best of our knowledge, a sigmoid optical neuron that generates the weighted optical summation of a WDM optical signal and utilizes an optical logistic sigmoid activation function unit. Our scheme employs a recently presented optical thresholding module [31] and a wavelength-encoded weighting scheme, with the optical neuron comprising four optical inputs that are weighted and summed producing a multi-level power signal before entering the sigmoid optical element. The sigmoid optical element is formed by a deeply saturated differentially-biased SOA-MZI followed by a SOA that performs as a Cross-gain Modulation-Wavelength Converter (XGM-WC), with preliminary results of its thresholding capabilities having been reported recently in [31]. We provide a detailed theoretical analysis of its sigmoid transfer function characteristics and validate experimentally the logistic sigmoid response, revealing a very good agreement between theory and experiment and almost excellent fitting to the mathematical expression of the logistic sigmoid function. Finally, the four-input optical sigmoid neuron is experimentally demonstrated in a four-pulse thresholding experiment, using 100ps long optical pulses and offering an up to 3.76dB Extinction Ratio (ER) improvement and a 100% improvement over power-level resolution capabilities in experimentally validated state-of-the-art photonic neuromorphic thresholding schemes [11, 12].

2. The photonic sigmoid activation unit: theory & experiment

A layout of a typical artificial neuron is depicted in Fig. 1(a). Each input X_n should be properly weighted by a certain weight W_n before all the signals enter the summation stage. Afterwards, the weighted summation is forwarded into the activation unit producing the neuron's output signal. Photonic counterparts that mimic this functionality have already been proposed in the literature [7, 14], with one of them being the WDM layout of Fig. 1(b). Here, each input $X_1 \dots X_n$ is imprinted onto a dedicated wavelength $\lambda_1 \dots \lambda_n$ via analog amplitude modulation, with the weights $W_1 \dots W_n$ being realized by Variable Optical Attenuators (VOAs). The weighted inputs are then multiplexed into a single optical branch, with the power level of the resulting WDM signal corresponding to the weighted summation of the n respective optical input signal power levels and used to trigger the optical activation function ϕ that completes the optical neuron.

The layout of the proposed photonic sigmoid activating device is presented in Fig. 2(a). A SOA-MZI interferometer operating in its deeply saturated regime is configured in a differentially-biased scheme [32] and is subsequently followed by a SOA that operates in its small-signal gain region, with both devices performing as wavelength converters. One continuous wave (CW) at λ_0 is driven to the input "C" of the SOA-MZI, with another CW at λ_1 being fed into SOA2 through the control arm "D" in order to realize the differentially-biased scheme, with both CW signals having high optical power levels and forcing both SOA1 and SOA2 to operate in their deeply saturated regime. An additional control pulse signal at λ_2 is attenuated by the bias attenuator in order to achieve the proper biasing of the activation function and then is split into 2 identical streams before being forwarded into the ports "A" and "H" of the SOA-MZI branches as co- and counter-propagating control beams. In this way, an inverse copy of the control signal imprinted on λ_0 is obtained at the switched output port "G" of the differentially-biased SOA-MZI, which is subsequently injected as control into the following SOA that operates as a XGM-WC and restores both the wavelength and the logic of the initial signal using an additional CW optical beam at λ_2 as its input signal.

The sigmoid transfer function of the photonic activation unit stems from the power equalization

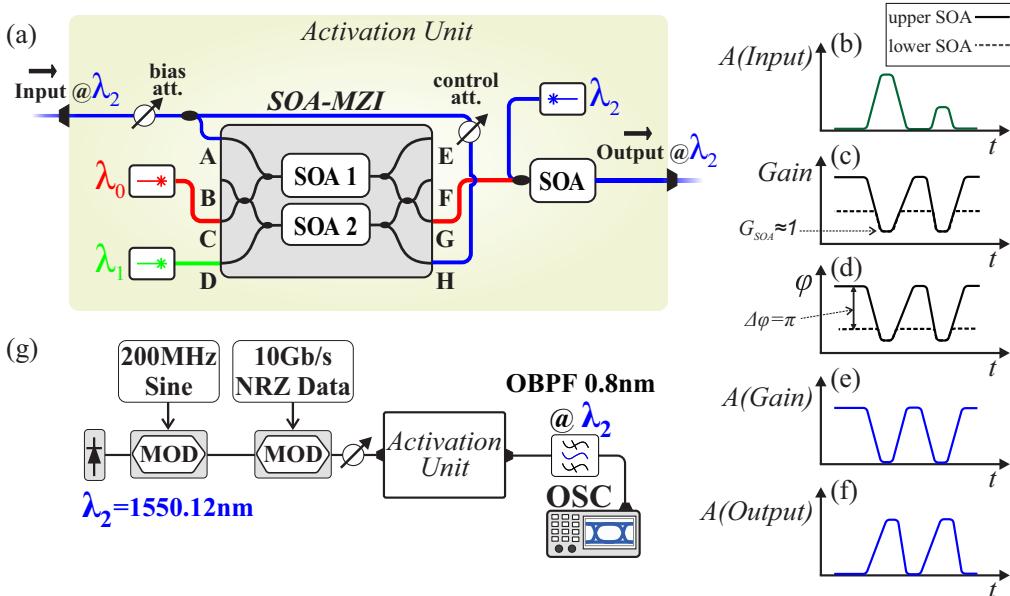


Fig. 2. (a) Photonic thresholder layout, (b)-(f) Principle of operation for the photonic thresholder, (g) Experimental setup for the measurement of transfer function.

properties of the deeply saturated differentially-biased SOA-MZI along with the nonlinear transfer function of the SOA XGM-WC operation. The injection of the appropriate amount of the λ_1 CW beam at SOA2 of the MZI leads the SOA2 gain close to its unitary end-point at the transparency region. At the same time, the CW input signal at λ_0 forces SOA1 to operate at a different gain level slightly above the transparency region, so that the differential gain between the 2 SOAs corresponds to a π phase-shift between the two SOA-MZI branches. Using this biasing scheme, the injection of a control pulse sequence with intense pulse peak power variation will result to the inverted copy of this signal at the SOA-MZI output but with almost power equalized pulses, as can be seen from the example described by Figs. 2(b)-(f). Figure 2(b) depicts two control pulses with different intensity levels that enter the SOA-MZI. Figure 2(c) illustrates the gain response of both the upper and lower SOA modules upon this control sequence is injected, clearly illustrating that every control pulse drives both SOA gains to their unitary end-point. Figure 2(d) shows the respective phase response experienced at both SOA-MZI branches, revealing that the initial π phase difference becomes zero during control pulse injection, following the principle described in [32]. Assuming that the lowest optical control pulse has enough power to drive the SOA gain to transparency, then obviously any optical control pulse will also yield to unitary SOA gain values, suggesting that optical control pulses with different intensity levels will yield the same $\Delta\phi = 0$ phase difference between the two SOA-MZI arms. This will result to a perfect "zero" level at the switched output port "G" of the SOA-MZI, with the respective "G"-port output sequence being depicted in Fig. 2(e). As can be observed, the "G"-port output sequence is an inverted but power equalized copy of the original control pulse sequence, which had, however, strongly intensity modulated pulses. Injecting this sequence into the subsequent SOA that operates as an XGM-WC, the signal gets back to its initial logic retaining its intensity equalized characteristics, as can be recognized in Fig. 2(f).

The experimental setup used for evaluating the transfer function characteristics of this layout is shown in Fig. 2(g). A CW at $\lambda_2 = 1550.12\text{nm}$ is injected into a Lithium Niobate (LiNbO_3) amplitude modulator driven by signal generator for producing a sinusoidal electric signal of 200MHz. The modulated signal is forwarded into the following LiNbO_3 modulator driven by a programmable Pulse Pattern Generator (PPG), realizing a periodic signal of 3 bits, with its pattern consisted of the "100" pulse sequence. As a result, the illustrated multi-level signal of Fig. 3(a) comprises 100psec pulses with their peak power levels forming a sinusoidal envelope that is properly attenuated by the respective attenuator before being forwarded as a control signal into the photonic activation unit. The optical signal exiting the activation function unit gets then filtered in an Optical Band-Pass Filter (OBPF) and is monitored at a digital oscilloscope. Figure 3(b) depicts the output pulse sequence when an average power level of 8dBm entered the activation unit, while Figs. 3(c), (d) illustrate the corresponding output sequences when power levels of 3 and -2dBm were injected as control signals, respectively.

The experimentally measured transfer function is illustrated by the black solid line in Fig. 3(e), (f) and was obtained by the following procedure: the peak power of every control pulse in Fig. 3(b)-(d) was calculated for every average control signal optical power level by scanning the average power level within a 16dB range using a step of 1dB. Taking into account that the captured time trace of the input signal (Fig. 3(a)) includes 25 pulses with different peak powers within a single period, this procedure has resulted to 400 different input peak power levels within the range of [-25, 0]dB. The transfer function shown in Figs. 3(e) and (f) in logarithmic and linear scale, respectively, has been calculated by measuring the ER value of every output pulse and correlating it to its corresponding input peak power level. This process yields an experimental curve that follows the average distribution of the ER values and has a statistical experimental y-error depicted via bars in both Figs. 3(e) and (f), with the highest y-error being 0.43dB. As can be seen in Fig. 3(e), the transfer function confines the output power levels between two discrete values formed by the 0dB and -11dB normalized ER value at the y-axis, corresponding to the "0"

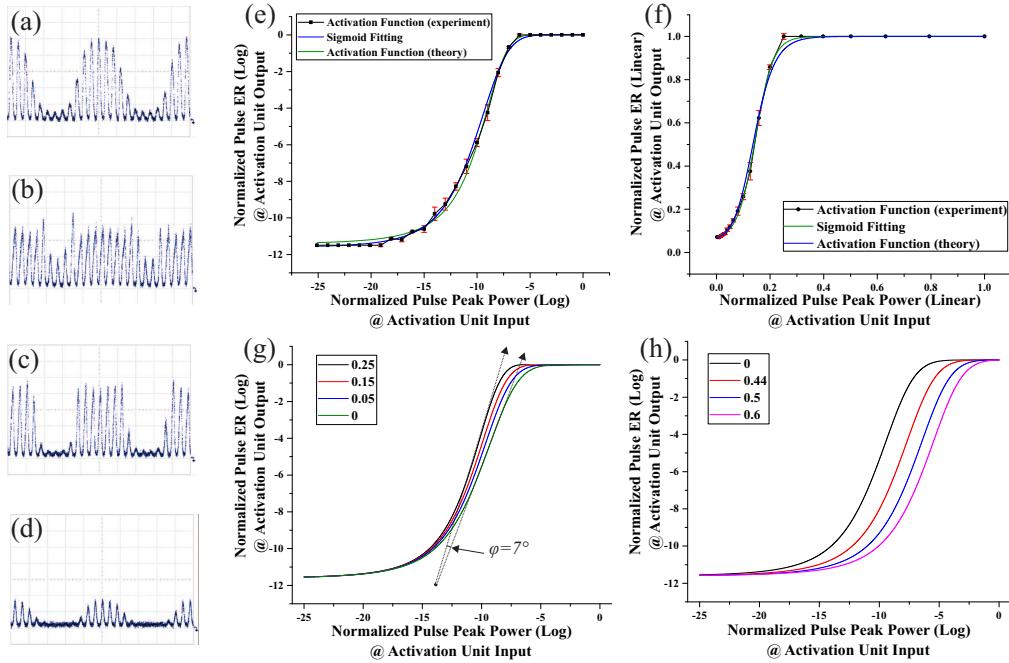


Fig. 3. (a) Multi-level signal for the activation function evaluation, (b)-(d) Time traces at activation's function output for 8, 3 and -2dBm average input optical power, Plot for the experimental transfer function and its y-error, the sigmoid fitting and the theoretical transfer function in: (e) logarithmic and (f) linear scale, Plot for the theoretical curve by varying the: (g) control attenuation factor af and (h) biasing attenuation factor b . y-axis scale: 3.5mV/div, x-axis scale: 1ns/div.

and "1" binary levels, respectively. It reveals a flat region around the "0" level, with the dynamic range of this flat region being 10.5dB. The transition between "0" and "1" has a range of 7.5dB, while the transfer function remains at a constant "1" level for a dynamic range of 7dB. In the linear plot of Fig. 3(f), the lower part of the sigmoid curve cannot extend below zero peak power levels at the x -axis since peak power levels can only be positive numbers. As such, the lower part flat region of the sigmoid shape can be hardly identified in this case, however it becomes evident when biasing the S-shape at higher x -axis input values, as will be shown in the next paragraphs.

The characteristics of the experimentally measured transfer function were analyzed by performing a nonlinear fitting with the generic form of the sigmoid function, which obeys to the following generic mathematical formula expressed by the Eq. (1):

$$f(x) = A_2 + \frac{A_1 - A_2}{\left(1 + e^{((x-x_0)/d)}\right)} \quad (1)$$

Figure 3(e) depicts the respective curve obtained when using $A_1 = 0.060$, $A_2 = 1.005$, $x_0 = 0.145$ and $d = 0.033$, revealing that the blue line is almost identical to the experimental transfer function.

The theoretical analysis of the proposed activation unit characteristics was carried out by relying on the established theoretical framework deployed in our previously published work [20, 32] for the SOA-MZI and SOA XGM-WC response. The optical power obtained at the switched and unswitched output ports of the SOA-MZI, denoted as "G" and "F" in the respective layout of Fig.

2(a), can be expressed by the Eqs. (2) and (3) below:

$$P_G(t) = P_{noise} + \frac{1}{4}P_{\lambda_0} \left[G_1(t) + G_2(t) - 2\sqrt{G_1(t)G_2(t)} \cdot \cos(\Delta\phi(t)) \right] \quad (2)$$

$$P_F(t) = P_{noise} + \frac{1}{4}P_{\lambda_0} \left[G_1(t) + G_2(t) + 2\sqrt{G_1(t)G_2(t)} \cdot \cos(\Delta\phi(t)) \right] \quad (3)$$

$$\Delta\phi(t) = -\frac{\alpha}{2} \ln[G_1(t)/G_2(t)] \quad (4)$$

where P_{noise} denotes the noise level of the system, P_{λ_0} the incoming CW optical input power at λ_0 , $G_1(t)$ and $G_2(t)$ stand for the SOA1 and SOA2 power gains, respectively, experienced by the CW input signal components, with α being the SOA's linewidth enhancement factor. During gain saturation, $G_1(t)$ and $G_2(t)$ gains are given by the followed time-dependent equations:

$$G_1(t) = \left[1 - \left(1 - \frac{1}{G_{CW(1)}} \exp \left(- \int_{-\infty}^t (1-b) \cdot P_{in}(t') dt' / U_{sat} \right) \right) \right]^{-1} \quad (5)$$

$$G_2(t) = \left[1 - \left(1 - \frac{1}{G_{CW(2)}} \exp \left(- \int_{-\infty}^t (1-af) \cdot (1-b) \cdot P_{in}(t') dt' / U_{sat} \right) \right) \right]^{-1} \quad (6)$$

where $G_{CW(1)}$ and $G_{CW(2)}$ denote the steady-signal gain of SOA1 and SOA2, respectively, in the absence of any control pulse, $P_{in}(t)$ refers to the peak power of the injected control pulse energy, U_{sat} is the saturation energy of the SOA and af denotes the attenuation factor applied to the control signal constituent injected into SOA2 and denoting the fraction of the control pulse energy injected into SOA2 compared to the corresponding control pulse energy injected into SOA1. The b parameter stands for the signal attenuation induced at the input signal prior being injected as the control signal into the SOA-MZI. The switched signal $P_G(t)$ exiting the SOA-MZI is fed then as the control signal into the SOA XGM-WC, where an input CW beam at λ_2 with an optical power of P_{λ_2} serves as the input signal. The equation that describes the output of SOA and consequently the output of the entire optical activation unit is expressed by the following mathematical formula:

$$P_{Activation\ Unit}(t) = P_{noise} + P_{\lambda_2} \left[1 - \left(1 - \frac{1}{G_{CW(SOA)}} \exp \left(- \int_{-\infty}^t P_G(t') dt' / U_{sat} \right) \right) \right]^{-1} \quad (7)$$

P_{noise} is equal to the system's noise level, $G_{CW(SOA)}$ stands for the SOA steady-state gain, and U_{sat} is again the SOA saturated energy level.

Equations (5), (6) and (7) show that the gains $G_1(t)$, $G_2(t)$ and the $P_{Activation\ Unit}$ saturate to a minimum value until the whole pulse energy has passed through the semiconductor, with the corresponding pulse energies being given by the integrals $\int_{-\infty}^t (1-b)P_{in}(t') dt' / U_{sat}$, $\int_{-\infty}^t (1-af)(1-b)P_{in}(t') dt' / U_{sat}$ and $\int_{-\infty}^t (1-b)P_G(t') dt' / U_{sat}$. From this time, that we denote hereinafter as t_s , the gain recovers back to its steady-state value until the next control pulse enters the SOAs, with the stimulated carrier lifetime constant τ_e [20]. In what follows, the bit period is assumed to be greater than the stimulated carrier recombination time, which allows us to ignore the gain recovery period and consider that G_1 , G_2 and the $P_{Activation\ Unit}$ values will return to their initial maximum values after a control pulse has been completely inserted into the semiconductors and before the arrival of a new control pulse. Replacing then Eqs. (4), (5) and (6) into (2), one gets the power $P_G(t)$ that exits the SOA-MZI through its "G" port proportionally to the control pulse peak power $P_{in}(t)$. Replacing the resulting expression into Eq. (7), the $P_{Activation\ Unit}$ is then expressed as a function of $P_{in}(t)$. Assuming Gaussian control pulses

with P_p being their peak power and $a(t)$ being their normalized Gaussian shape waveform, so that $P_{in}(t) = P_p \cdot a(t)$, the time integral of $P_{in}(t)$ concludes to the expression $P_p \int_{-\infty}^t a(t')dt'$. As we are interested in investigating the peak power of the $P_{Activation\ Unit}$ waveform for extracting its output versus input power transfer function characteristics, the maximum value of $P_{in}(t)$ for every injected control pulse coincides temporally with the time t_s where the whole pulse has been just inserted into the semiconductors. This condition allows for the replacement of the time-integral $\int_{-\infty}^t a(t')dt'$ with a time-independent, constant value A that represents the total area covered by the pulse waveform. This value corresponds to the limit of the integral at infinity. Hence, the $P_{in}(t)$ can be expressed as $P_{in}(t) = P_p \cdot A$, where $A = \int_{-\infty}^t a(t')dt'$ and finally the $P_{Activation\ Unit}$ is obtained as a function of the control pulse peak power P_p . The theoretically obtained peak power values of $P_{Activation\ Unit}(t)$ at $t = t_s$ resulting from this mathematical analysis has been plotted for a broad range of P_p values and is depicted in Fig. 3(e) as the transfer function of the with a solid green line, using a noise constant of $P_{noise} = 0.0064mW$, SOA gains equal to $G_{CW1} = 2.85$, $G_{CW2} = 1.42$, and $G_{CW(SOA)} = 1000$, a SOA linewidth enhancement factor $\alpha = 6$, a saturation energy level of $U_{sat} = 1000fJ$ assuming the same value for all three SOAs, an optical power for the input CW of SOA-MZI and the SOA of $P_{CW} = 0.5mW$ and $P_{CW2} = 0.015mW$, respectively, while the attenuation and bias factor were equal to $af = 0.05$ and $b = 0$, respectively. As can be observed in Figs. 3(e), (f), the resulting graph follows closely both the experimentally measured transfer function as well as the sigmoid fitting, providing a solid theoretical basis for the optical sigmoid activation unit. Having confirmed the capability of our theoretical analysis in validating the experimental results, we can exploit the theoretical transfer function towards analyzing how important characteristics of the sigmoid curve can be affected by certain device parameters. Figure 3(g) illustrates how the slope of the theoretically obtained curve can be adjusted by varying the af variable, where a slope variation of $\phi = 7^\circ$ is shown for an af range between $[0, 0.25]$. Figure 3(h) reveals that the theoretical curve can also shift to higher input peak power values along the x -axis when changing the bias attenuation factor b , depicting a shift of close to $4.5dB$ when varying b between $[0, 0.6]$. This indicates that this attenuation factor has the role of the transfer function bias, similar to the role that x_0 performs in the typical sigmoid equation shown in Eq. (1).

3. Experimental demonstration of the all-optical sigmoid neuron

Following the experimental and theoretical validation of the transfer function characteristics, the sigmoid activation element has been incorporated in a complete optical neuron to evaluate its performance with weighted and summed optical signals. Figure 4 depicts the layout of the experimental setup that is used for the evaluation of the WDM-enabled optical neuron when employing the optical sigmoid activation unit described in Section 2. Four CWs at $\lambda_0 = 1549.4nm$, $\lambda_1 = 1550.2nm$, $\lambda_2 = 1551.0nm$, and $\lambda_3 = 1551.8nm$ are forwarded into 4 respective LiNbO₃ modulators, with each of them being driven by a PPG to produce a periodic signal with 100psec long pulses separated by 200psec time-spacing intervals. More specifically, in a frame of 1.6nsec, 4 pulses at λ_0 were produced by the first modulator, 3 pulses at λ_1 by the second modulator, 2 pulses at λ_2 by the third and 1 pulse at λ_3 by the fourth modulator as well. Every signal was multiplied by a respective weight implemented by a corresponding VOA before getting the proper time-synchronization via respective Optical Delay Lines (ODLs) and multiplexed in a subsequent Arrayed Waveguide Grating (AWG) multiplexer to produce a multi-level WDM signal obtained at the AWG output. The visualization of the corresponding time delay of each weighted input and the multi-level WDM signal at the AWG output are depicted in the corresponding insets of Fig. 4. After exiting the AWG, the multi-level WDM signal enters the VOA of the activation unit that is used to adjust the optical power of the weighted WDM signal, defining in this way the threshold and serving as the biasing factor. A 3dB coupler splits the optical signal into 2 identical streams, with the first one being injected into port A of the SOA-MZI. To realize the differentially-biased

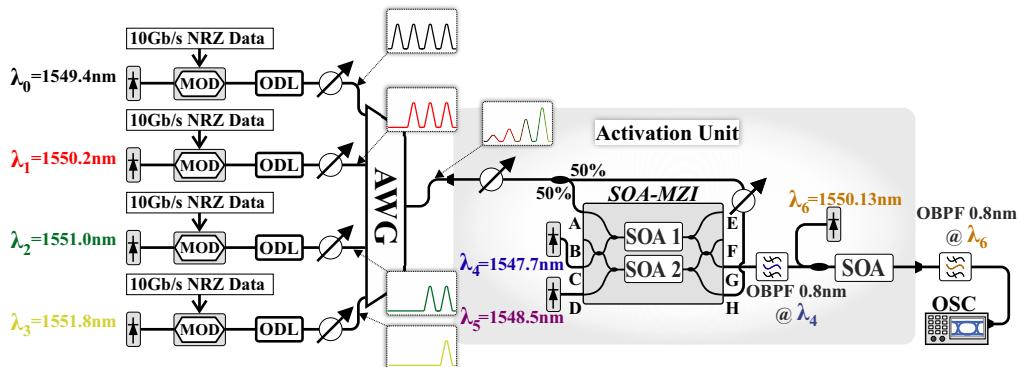


Fig. 4. Experimental setup for the optical sigmoid neuron.

scheme, the second stream needs to be attenuated by a VOA before being injected into the port "H" of the SOA-MZI. For completing this scheme, two CWs at $\lambda_4 = 1547.7\text{nm}$ $\lambda_5 = 1548.5\text{nm}$ are injected into the ports "C" and "D". The wavelength converted signal emerging at the "G"-port of the SOA-MZI is filtered by an OBPF at $\lambda_4 = 1547.7\text{nm}$ and is then fed as the control signal into a SOA along with a CW input signal at $\lambda_6 = 1550.13\text{nm}$. In this way, the SOA operates as an XGM wavelength converting gate and the output signal is filtered by an OBPF at $\lambda_6 = 1550.13\text{nm}$ before it reaches the oscilloscope for signal quality evaluation purposes.

Time traces and eye diagrams that were obtained during the evaluation of the sigmoid optical neuron are depicted in Fig. 5. Figure 5(a) illustrates the pulse sequence and the corresponding eye diagram of the multi-level WDM signal at the input of the activation unit, where the first pulse has an ER of 0.54dB, the second 3.01dB, the third 5.3dB and the last one 8.66dB. Injecting 11.5dBm and 10.7dBm of this signal into the SOA-MZI's ports "A" and "H", respectively, sets the threshold of the sigmoid activation function at the lowest power level, denoted as "*level 1*" in Fig. 5(a) with all the four power levels being above this threshold value. The time trace at the "G" output port of the SOA-MZI is depicted in Fig. 5(b) with every inverted pulse being at the zero level, while the time trace and the respective eye for the output of the neuron for this threshold level are shown in Fig. 5(c), clearly confirming that all 4 incoming pulses that exceed the threshold "*level 1*" have been transformed to respective power equalized output pulses with an amplitude modulation index (AM) of 4.4dB and an minimum ER of 3dB. The AM is here defined as the ratio of the peak power level of the highest to the weakest pulse, while the ER is defined as the ratio of the peak power level of the weakest pulse to the average power level of zero. This suggests that the minimum ER between the one and the zero power level has been improved by 3.76dB compared to the 0.54dB ER value of the original multi-level WDM input signal. As can be seen in the time trace of Fig. 5(d), the first pulse of the SOA-MZI output is above the zero level when the biasing threshold of activation unit is configured at "*level 2*" by means of increasing the bias VOA-induced losses and reducing the input signal optical power by 3.4dB. Figure 5(e) illustrates the time trace and eye diagram at the output of the optical neuron showing that the weakest input optical pulse that was below the threshold has disappeared, while the three pulses that were above the threshold are retained at the output and have turned into a power equalized pulse sequence with an AM of 0.7dB and a minimum ER of 5dB. This indicates that the input signal ER of 2.6dB, defined as the ratio between the peak power of the weakest pulse that lies above threshold to the peak power of the highest pulse that is below threshold, has been transformed into a 5dB ER at the threshold output. Reducing further the input signal optical power of the activation unit by 3.3dB by means of the bias VOA stage, the threshold is set at "*level 3*", forcing the first and the second pulse to provide inverted pulses above the zero level

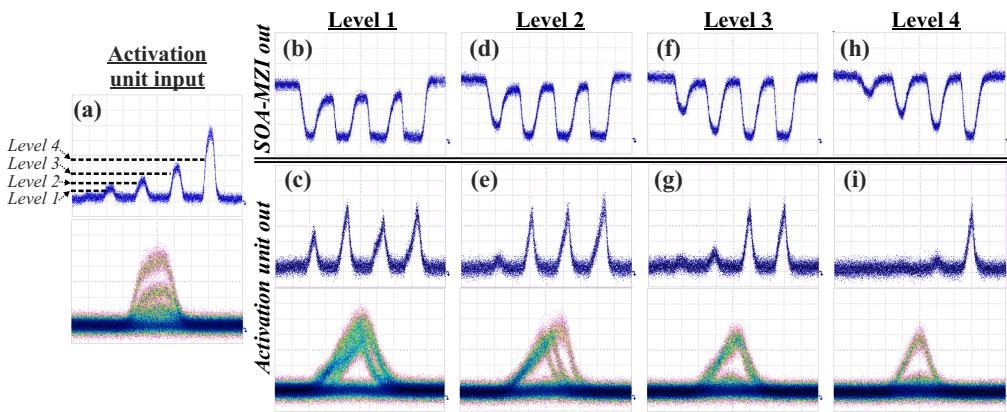


Fig. 5. (a) Time trace (200psec/div) of the multi-level signal at the input of activation unit, where the dashed lines denote the different power levels. The corresponding eye diagram is also presented below (50psec/div). (b), (d), (f) and (h): Time traces at the output of the SOA-MZI, (c), (e), (g), and (i): Time traces (200psec/div) and eye diagrams (50psec/div) at the output of the optical neuron. y-axis scale: (4mV/div).

at the output of SOA-MZI, as illustrated in Fig. 5(f). Hence, the sigmoid optical neuron output retains only the two higher-level pulses suppressing the two pulses being below the threshold level, as shown by the time trace and eye diagram of Fig. 5(g). AM and minimum ER of the recorded signal in this case were calculated to be 0.65dB and 4.1dB, respectively, suggesting an ER increase of 1.5dB compared to the input ER value of 2.6dB when this is defined as the ratio between the peak powers of the weakest pulse above the threshold and the strongest pulse below the threshold. Finally, configuring the threshold at "level 4" has been done by reducing the optical power at the input of activation unit by 3.9dB in addition, resulting to a SOA-MZI output of Fig. 5(h) where only the inverted version of the last pulse reaches the zero level, and a neuron output that comprises a single optical pulse with an ER of 4.8dB, as shown in Fig. 5(i). In this case, the input signal ER of 3.5dB increases by 1.3dB at the threshold output, confirming again the improved signal quality delivered by the activation unit to the output pulse sequence. During the evaluation of the proposed sigmoid optical neuron the operational conditions were as follows: SOA-MZI had its two SOAs SOA1 and SOA2 driven by a dc current of 240mA and 280mA, respectively. For the ports "C" and "D", the CW optical power levels were 5.5dBm and 6.5dBm, while the highest optical power for the control signals of port "A" and "H" was 11.5dBm and 10.7 dBm, respectively. For the output SOA, the driving current was set at 300mA and the input CW signal had an optical power level of -10.5dBm.

4. Conclusion

We have presented for the first time a sigmoid all-optical artificial neuron based on a WDM input & weighting scheme and a novel sigmoid activation function, realized by a deeply saturated differentially-biased SOA-MZI followed by a SOA that operates in its small-signal gain regime. The theoretical study of the activation unit has been presented and verified experimentally, revealing an excellent fit with the logistic sigmoid activation function and enabling in this way the direct employment of its mathematical expression in neural network training and execution processes. Finally, the evaluation of the proposed optical neuron was realized via the experimental demonstration of thresholding in a WDM weighted and summed optical signal with 4 different power levels and 100psec long pulses, offering a 100% improvement in the number of threshold level resolutions compared to experimentally demonstrated state-of-the-art neuromorphic optical

activation elements [11, 12] and revealing an ER improvement of up to 3.76dB.

Funding

EU H2020 projects ICT-STREAMS (688172) and L3MATRIX (688544).

References

1. J. M. Shalf and R. Leland, "Computing beyond Moore's Law," *Computer* **48**, 14–23 (2015).
2. F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G. J. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, and D. S. Modha, "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," *IEEE Transactions on Comput. Des. Integr. Circuits Syst.* **34**, 1537–1557 (2015).
3. S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, and A. D. Brown, "Overview of the SpiNNaker system architecture," *IEEE Transactions on Comput.* **62**, 2454–2467 (2013).
4. N. P. Jouppi, A. Borchers, R. Boyle, P.-I. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, C. Young, T. V. Ghaemmaghami, R. Gottipati, W. Gulland, R. Hagmann, C. R. Ho, D. Hogberg, J. Hu, R. Hundt, D. Hurt, J. Ibarz, N. Patil, A. Jaffey, A. Jaworski, A. Kaplan, H. Khatan, D. Killebrew, A. Koch, N. Kumar, S. Lacy, J. Laudon, J. Law, D. Patterson, D. Le, C. Leary, Z. Liu, K. Lucke, A. Lundin, G. MacKean, A. Maggiore, M. Mahony, K. Miller, R. Nagarajan, G. Agrawal, R. Narayanaswami, R. Ni, K. Nix, T. Norrie, M. Omernick, N. Penukonda, A. Phelps, J. Ross, M. Ross, A. Salek, R. Bajwa, E. Samadiani, C. Severn, G. Sizikov, M. Snelham, J. Souter, D. Steinberg, A. Swing, M. Tan, G. Thorson, B. Tian, S. Bates, H. Toma, E. Tuttle, V. Vasudevan, R. Walter, W. Wang, E. Wilcox, D. H. Yoon, S. Bhatia, and N. Boden, "In-Datacenter Performance Analysis of a Tensor Processing Unit," *ACM SIGARCH Comput. Archit. News* **45**, 1–12 (2017).
5. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics* **11**, 441–446 (2017).
6. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science* **361**, 1004–1008 (2018).
7. A. N. Tait, J. Chang, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, "Demonstration of WDM weighted addition for principal component analysis," *Opt. Express* **23**, 12758 (2015).
8. A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing," *J. Light. Technol.* **32**, 4029–4041 (2014).
9. A. N. Tait, T. F. De Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Reports* **7**, 1–10 (2017).
10. C. Mesaritakis, A. Kapsalis, A. Bogris, and D. Syvridis, "Artificial Neuron Based on Integrated Semiconductor Quantum Dot Mode-Locked Lasers," *Sci. Reports* **6**, 1–10 (2016).
11. D. Rosenbluth, K. Kravtsov, M. P. Fok, and P. R. Prucnal, "A high performance photonic pulse processing device," *Opt. Express* **17**, 22767 (2009).
12. K. S. Kravtsov, M. P. Fok, P. R. Prucnal, and D. Rosenbluth, "Ultrafast All-Optical Implementation of a Leaky Integrate-and-Fire Neuron," *Opt. Express* **19**, 2133 (2011).
13. H. T. Peng, M. A. Nahmias, T. F. De Lima, A. N. Tait, B. J. Shastri, and P. R. Prucnal, "Neuromorphic Photonic Integrated Circuits," *IEEE J. Sel. Top. Quantum Electron.* **24**, 1–16 (2018).
14. I. Chakraborty, G. Saha, A. Sengupta, and K. Roy, "Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons," *Sci. Reports* **8**, 1–10 (2018).
15. M. A. Nahmias, A. N. Tait, L. Tolias, M. P. Chang, T. Ferreira De Lima, B. J. Shastri, and P. R. Prucnal, "An integrated analog O/E/O link for multi-channel laser neurons," *Appl. Phys. Lett.* **108** (2016).
16. M. A. Nahmias, B. J. Shastri, A. N. Tait, and P. R. Prucnal, "A leaky integrate-and-fire laser neuron for ultrafast cognitive computing," *IEEE J. on Sel. Top. Quantum Electron.* **19**, 1–12 (2013).
17. A. N. Tait, B. J. Shastri, M. P. Fok, M. A. Nahmias, and P. R. Prucnal, "The DREAM: An integrated photonic threshold," *J. Light. Technol.* **31**, 1263–1272 (2013).
18. N. S. Rafidi, K. S. Kravtsov, Yue Tian, M. P. Fok, M. A. Nahmias, A. N. Tait, and P. R. Prucnal, "Power Transfer Function Tailoring in a Highly Ge-Doped Nonlinear Interferometer-Based All-Optical Threshold," *IEEE Photonics J.* **4**, 528–534 (2012).
19. G. Morthier, J. Sun, T. Gyselings, and R. Baets, "A novel optical decision circuit based on a Mach-Zehnder or Michelson interferometer and gain-clamped semiconductor optical amplifiers," *IEEE Photonics Technol. Lett.* **10**, 1162–1164 (1998).
20. N. Pleros, C. Bintjas, G. T. Kanellos, K. Vlachos, H. Avramopoulos, and G. Guekos, "Recipe for intensity modulation reduction in SOA-based interferometric switches," *J. Light. Technol.* **22**, 2834–2841 (2004).
21. J. George, R. Amin, A. Mehrabian, J. Khurgin, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks," in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, (Optical Society of America, 2018), p. SpW4G.3.
22. X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," (2011), pp. 315–323.
23. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, (2015), pp. 1026–1034.

24. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation* **9**, 1735–1780 (1997).
25. N. Passalis, G. Mourgas-alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, "Variance preserving initialization for training deep neuromorphic photonic networks with sinusoidal activations," submitted to ICASSP (2019).
26. K. Gregor, I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra, "Draw: A recurrent neural network for image generation," arXiv preprint arXiv:1502.04623 (2015).
27. K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," *Sci. Total. Environ.* **572**, 169–176 (2015).
28. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, (2017), pp. 5998–6008.
29. Y. Kim, C. Denton, L. Hoang, and A. M. Rush, "Structured attention networks," arXiv preprint arXiv:1702.00887 (2017).
30. M. Miscuglio, A. Mehrabian, Z. Hu, S. I. Azzam, J. George, A. V. Kildishev, M. Pelton, and V. J. Sorger, "All-optical nonlinear activation function for photonic neural networks [Invited]," *Opt. Mater. Express* **8**, 3851 (2018).
31. G. Mourgas-alexandris, A. Tsakyridis, T. Alexoudi, K. Vrysokinos, and N. Pleros, "Optical thresholding device with a sigmoidal transfer function," in *Proceedings Of Photonics in Switching and Computing*, (2018).
32. M. Spyropoulou, N. Pleros, K. Vrysokinos, D. Apostolopoulos, M. Bougioukos, D. Petrantonakis, A. Miliou, and H. Avramopoulos, "40 Gb/s NRZ wavelength conversion using a differentially-biased SOA-MZI: Theory and experiment," *J. Light. Technol.* **29**, 1489–1499 (2011).