Photonic Neuron with on Frequency-Domain ReLU Activation Function

Margareta Vania Stephanie, *Student Member, IEEE*, Lam Pham, Alexander Schindler, Tibor Grasser, *Fellow, IEEE,* Michael Waltl, *Senior Member, IEEE*, and Bernhard Schrenk, *Member, IEEE*

***Abstract*—Driven by an exponential growth of data, neuromorphic computing has risen in popularity as a new method for high-performance computing. The adopted neural network (NN) model relies on parallel processing between neurons and synapses, which reduces the energy consumption and boosts the computational efficiency. Photonics empowers neuromorphic processors through its inherent parallelism, along with high speed and unique bandwidth characteristics. Yet, it requires to transfer each constituent of the NN model to the optical realm, including the challenging nonlinear part of an activation function. Towards this direction, we experimentally demonstrate a photonic rectified linear unit (ReLU) function by employing frequency coding of neural signals in combination with a periodic optical filter. Furthermore, we show that multiple neural sub-circuits can be collapsed over the proposed photonic ReLU hardware and further evaluate the possibility to integrate weighting functionality with the frequency-domain ReLU as a way to further simplify the optical NN. For these demonstrations, we accomplish a low penalty of 1-3% in terms of accuracy when transferring the Iris flower classification challenge from the digital to the optical realm. Finally, we introduce an efficient translucent interface between the linear and nonlinear circuits of a photonic neuron, utilizing an optical frequency-coder that is directly driven by the photocurrent of a preceding photodetector – without the need for electrical amplification.**

***Index Terms*—Multilayer perceptrons, Neural network hardware, Optical signal processing, Neuromorphic photonics**

I. Introduction

DIGITAL electronic computing has been developed over decades, eventually reaching the scale of high-performance supercomputers [1-2]. These processing machines are based on the von-Neumann architecture, where communication between a dedicated central processing unit (CPU) and a physically separated memory is occurring sequentially [3]. This specific computing architecture is currently hitting a bottleneck due to the limited data transfer rate between the high-speed processor and large memory pages, resulting in a latency obstacle that is causing the overall computing efficiency to slow down [4-5]. Additionally, dense and long electronic interconnects lead to a limited signal bandwidth, heat dissipation, and high energy consumption [6].

On the contrary, the human brain hosts about 10^{11} neurons; each of these is connected to ~10^4 inputs, resulting in a total of 10^{15} synaptic connections [7]. More interestingly, the nature-built brain facilitates this computation interconnect with a rather low power consumption of only 20 W. This corresponds to a computational efficiency that is eight orders-of-magnitude higher than human-created digital machines [8]. Due to this astonishing advantage over the classical computing architecture, a bio-inspired neural network (NN) model has been an integral part of modern technologies concerning artificial intelligence (AI). Over the past few years, successful AI software-implementations leveraging on classical computer hardware have triggered the development of neuromorphic-based computing architectures that aim at an energy-efficient information processing machine. In this architecture, both CPU and memory are governed by the neurons and synapses [9]. Owing to parallelism, the delay in data processing can be minimized and high-speed information processing at low power consumption can be unlocked for practical applications on the longer term [10].

Photonic is a suitable building block in turning the notion of neuromorphic computing into reality [11-12]. It drives NN processing, thanks to its inherent abilities to efficiently multiplex signals (parallelism) and its low power consumption in view of large electro-optic bandwidths [13]. Optical NN (ONN) based computing machines have been conceptually proven to offer low latency and fast signal processing in the GHz range for a wide range of applications [14-17]. The fundamental NN architecture comprises a multiply-accumulate (MAC) operation incorporating a weighting process, followed by a nonlinear activation function. Linearity is at the heart of optics; therefore, weighted summation has been demonstrated for a plethora of approaches, including Mach-Zehnder interferometer (MZI) meshes [18], micro-ring resonator (MRR) weight banks [19], semiconductor optical amplifier (SOA) based variable-gain networks [20], and frequency-coded synapses in combination with spectral processing [21].

In addition, the activation element in a NN plays the important role of discriminating data using a decision boundary, involving functions such as the rectified linear unit (ReLU), the sigmoid or hyperbolic tangent functions – each of

Manuscript received January 22, 2024. This work was supported by the Austrian Research Promotion Agency FFG through the JOLLYBEE project (grant No. 887467).

M. V. Stephanie, L. Pham, A. Schindler, and B. Schrenk are with the AIT Austrian Institute of Technology, Center for Digital Safety&Security, Giefinggasse 4, 1210 Vienna, Austria (e-mail: margareta.stephanie@ait.ac.at).

T. Grasser and M. Waltl are with the Institute for Microelectronics, TU Wien, Gusshausstrasse 27-29, 1040 Vienna, Austria (e-mail: grasser@iue.tuwien.ac.at; waltl@iue.tuwien.ac.at).

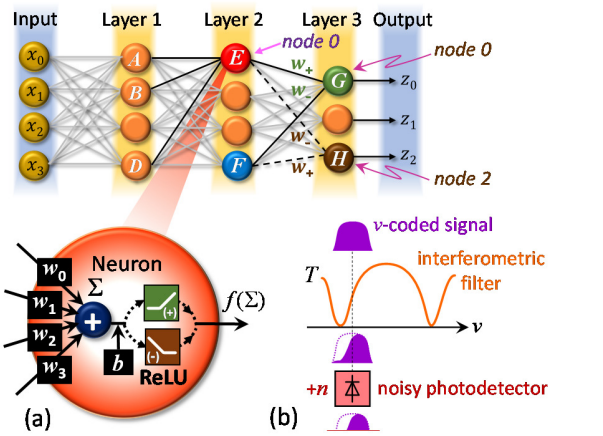

Fig. 1. Neural network based on the multilayer perceptron scheme, which is considered in the offline training as a benchmark for the optical neural network evaluation. (a) A functional diagram of a photonic neuron, which fundamentally comprises a linear weighted summation Σ and a ReLU activation function. (b) Photonic ReLU concept, employing an interferometric filter slope with a noisy photodetector to realize a rectification operation.

these being characterized through different advantages and suited for a specific set of applications [22]. However, nonlinear operation in view of either of those functions is rather challenging to accomplish in the realm of photonics. In the past few years, research efforts on optical activation functions have demonstrated the feasibility to practically realize opto-electronic or purely optical representations of these NN elements. One proposal relies on the use of electro-absorption modulators (EAM) to realize an electro-optic sigmoid activation [23]. A similar proposal can be obtained by using other modulators such as MRR, MZI on-chip with different architectures [24, 25, 26], or its combination with optoelectronic detectors [27]. The applied phase shift during modulation is the parameter that modifies the profile of the activation function, including the ReLU-like function. However, these electro-optic schemes require preceding opto-electronic conversion circuits.

An alternative all-optical approach relies on an MZI-integrated SOA configuration, which is operated in the deeply saturated differentially biased regime, followed by another SOA that performs a cross-gain modulation-based wavelength converter, showing a sigmoid-like transfer function [28]. However, the device occupies a relatively large footprint, which hinders the scaling for a larger-scale NN integration. Similar approaches have been shown by implementing SOA-based wavelength converters in which the synaptic weight is governed by the variation of SOA gain and also realizing a sigmoid function [29], or by employing a self-induced polarization rotation effect in a single SOA to achieve an arbitrary activation function such as ReLU or Softplus [30]. Furthermore, it is also possible to employ a plasma effect by changing the free carrier concentration in a waveguide [31] to vary the coupling ratio of a Mach-Zehnder coupler to eventually detune the wavelength between MRR resonance and the input signals. By tuning the thermal heaters that influence the amplitude and phase biases of these devices, a variety of nonlinear activation functions can be generated such

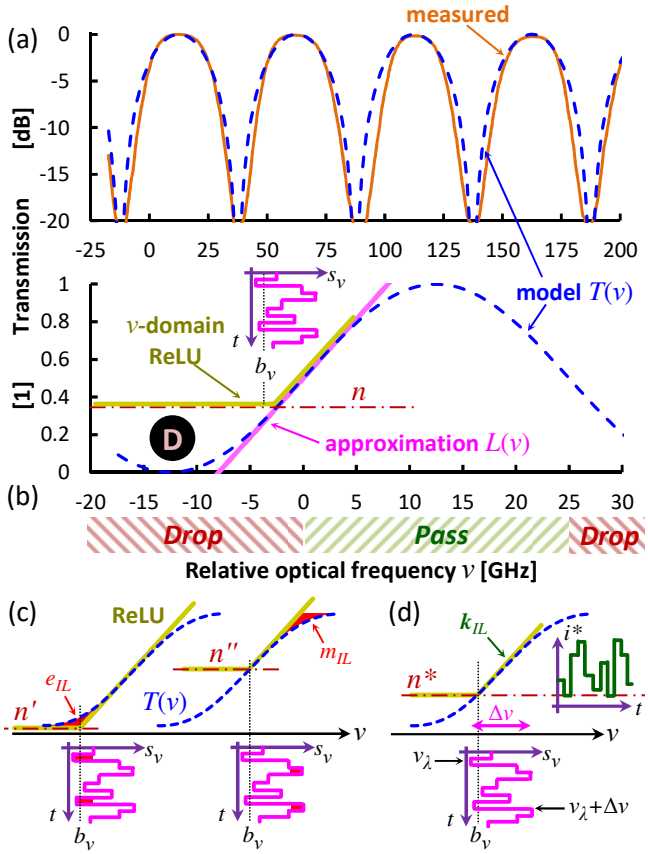

Fig. 2. (a) Measured and modelled $T(\nu)$ transmission function of an optical 25/50 GHz IL and (b) linear approximation $L(\nu)$ around its quadrature point, which defines the drop and pass region of the IL. Signal clipping of the neural data s_ν is obtained through the noise background n of a photodetector, introducing a drop window (D). (c) Sub-optimal noise thresholds where non-linear distortion for the neural signal appears at the extinction e_{IL} and maximum transmission m_{IL} of the IL. (d) Optimal threshold n^* yielding a rectified photocurrent i^* defined by the frequency deviation $\Delta\nu$, the ReLU bias b_ν and the frequency-to-intensity conversion k_{IL} of the IL.

as sigmoid, radial-basis and clamped ReLU. Although this scheme enables the reconfigurability of arbitrary functions, it is presently bound to latency and large energy consumption. Other MRR-based activation functions have been realized via silicon MRR loaded with phase change materials [32] or in a Ge/Si hybrid structure [33], showing different types of ReLU. However, these approaches necessitate very specific fabrication processes.

Further studies have shown that a semiconductor laser can be used to generate ReLU-like function by using an integrated III-V semiconductor membrane laser [34] or by exploiting the bistability of an injection-locked Fabry–Perot semiconductor laser [35]. Nonetheless, these schemes still require external modulation, which complicated their overall integration. Compelling results have been shown using a nonlinear periodically-poled thin-film Lithium-Niobate waveguide [36], exploiting second harmonic generation and degenerate optical parametric amplification as resources to realize a ReLU-like function. A linear rectification has been achieved by choosing the bias pulse power and the waveguide length, but the use of a rather exotic material and optical pumps might likely make it difficult to scale up this NN concept.

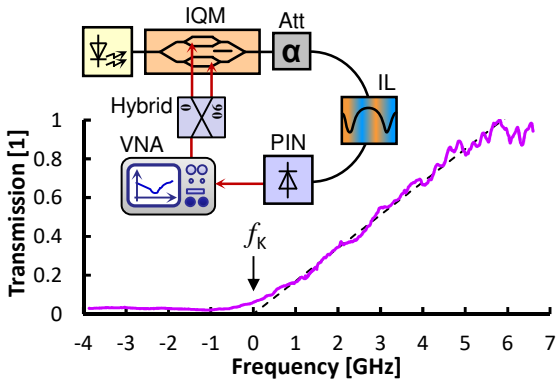

Fig. 3. Experimental setup to characterize the response of the photonic ReLU function and resulting transmission function after clipped photodetection.

In this work, we investigate another approach of ONN scheme based on optical frequency coding, as an extension of our previous findings [37, 38]. As a first step, we characterise and evaluate an experimental implementation for a photonic ReLU activation function via frequency-coded signals by employing a combination of chirped directly modulated laser (CML) and an optical interleaver (IL) as a spectral processing function. We achieve a penalty of 3% in accuracy with respect to an all-digital NN (DNN), which provides an inherent accuracy of 93% when being tasked with Iris flower classification. We then include the proposed ReLU as part of a photonic neuron, together with the linear weighted summation. We found a small 1% penalty for this hybrid ONN implementation by sharing the photonic ReLU hardware within multiple neural sub-circuits. Furthermore, the IL-based ReLU builds on elements that are well known from the field of optical telecommunications and can be further combined with the weight polarity assignment for neural signals. This promises a simplification of the overall photonic neuron and thus a smoother scaling of an ONN.

The paper is organized as follows. Section II describes the notions behind the photonic ReLU operation based on synaptic frequency coding and introduces the application framework. Section III evaluates the proposed ReLU concept in comparison with an electro-optic ReLU scheme. Section IV expands this characterisation with simultaneous weighting operation and shows that neural sub-circuits can be collapsed over the same optical hardware. Section V introduces a novel direct-drive scheme that permits an efficient translucent layer-to-layer concatenation within the ONN. Finally, Section VI draws a conclusion on the presented findings.

II. Linear Rectification through Frequency Coding

A. Frequency-Domain ReLU Function

Figure 1(a) introduces a photonic neuron within the ONN. It has a linear part, devoted to a weighted summation, and a nonlinear part responsible for neural activation. This second aspect is supported through a photonic ReLU function, which is implemented according to Fig. 1(b). Here, the optical input signal to the ReLU is coded in optical frequency ν, yielding a neural signal representation according to

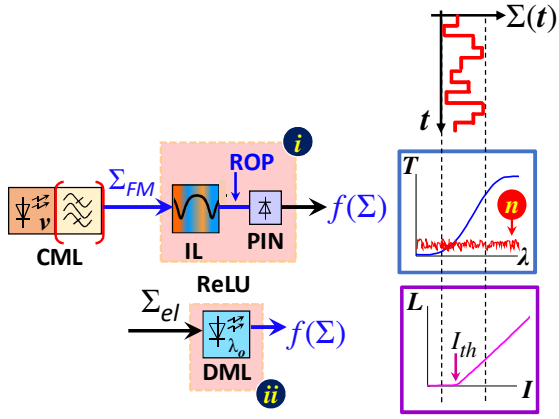

Fig. 4. Practical configuration and principal transmission characteristics of photonic ReLU schemes using (*i*) a combination of chirped-laser CML, interleaver and noisy PIN photodetection, which together with a low received optical power (ROP) is responsible for the noise (*n*) that clips the signal, and (*ii*) a directly modulated laser (DML).

$$s_\nu(\nu) = \nu_\lambda + \Delta\nu\, s(t), \tag{1}$$

where $s(t)$ is the neural data, ν_λ is the optical frequency of the frequency-coded signal and $\Delta\nu$ its peak-to-peak frequency deviation applied during coding. The signal s_ν then passes through the quadrature point of an interferometric filter, which for example can be an optical IL as it is commonly used for optical communication systems. We can express the transmission of such an IL through the periodic function

$$T(\nu) = \frac{1}{2}\left[1 - \cos\left(2\pi\,\delta T\,(\nu - \nu_o)\right)\right], \tag{2}$$

where ν_0 is the offset of the transfer function to the ITU-T grid, and δT is the delay parameter that defines the free spectral range and thus the grid spacing of the IL. Figure 2(a) presents the measured transmission of a 25/50 GHz IL, together with its modelled transmission T. This transfer function for frequency-to-intensity conversion will determine the activation slope of the ReLU function.

Next, we can approximate the transmission function at its quadrature point at T = -3dB, which yields the linear function

$$L(\nu) = \alpha + \beta\left[\nu - \nu(T=0)\right], \tag{3}$$

for which

$$\alpha = T\left(\nu(T=0)\right) = \frac{1}{2}, \quad \beta = \frac{d}{d\nu}T\left(\nu(T=0)\right) = \pi\,\delta T \tag{4}$$

This linear approximation is introduced to Fig. 2(b) together with the model T of the IL. This IL function passes the spectral components of the neural data signal s_ν in a way that – in case of the exemplary spectral allocation shown in Fig. 1(b) – lower frequencies would be suppressed yet not rectified. To accomplish the latter, the linear function L acts in combination with the added noise n of the subsequent photodetector, which covers the drop window (D) with a noise background, thus

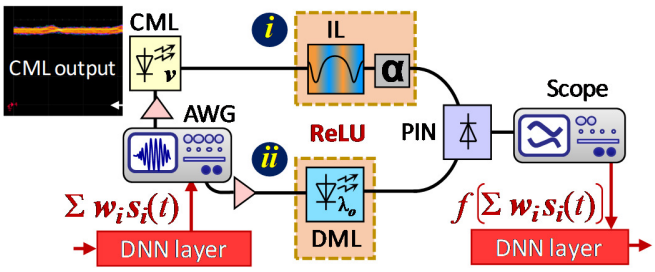

Fig. 5. Experimental setup to evaluate the performance of ReLU units based on (*i*) an optical interleaver and a PIN photodetector, in combination with a CML emitter, and (*ii*) an electro-optic approach by employing a DML.

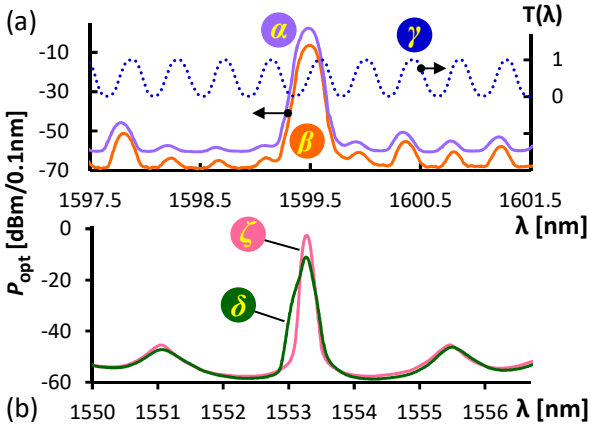

Fig. 6. (a) Optical spectra of CML+IL based ReLU implementation before (α) and after (β) the IL, including the IL transmission γ. (b) DML output spectrum for bias points above (ζ) and below (δ) the threshold current I_{th}.

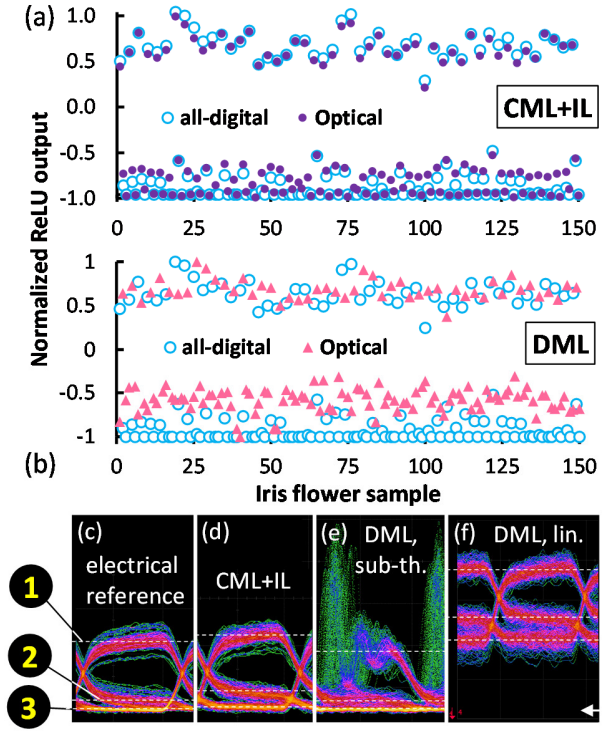

Fig. 7. Normalized ReLU output from all-digital NN (○) and photonic ReLU implementations based on (a) CML+IL (●), and (b) a DML (▲) for 150 Iris flower samples. The eye diagram on the bottom shows the output ReLU based on (c) all-digital NN, (d) photonic ReLU using CML+IL, (e) photonic ReLU using DML with sub-threshold current, and (f) above the threshold in the linear regime.

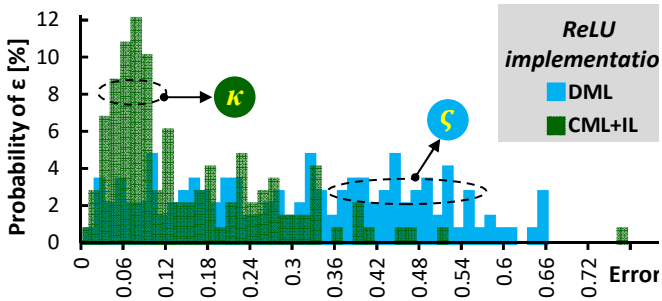

Fig. 8. Error histogram of photonic ReLU operation based on CML+IL (κ), which is lower than using DML implementation (ς).

clipping the neural data signal as defined by the bias b_v for the spectral alignment of s_v. This eventually provides the ReLU function required for the ONN implementation.

The clipping threshold n will generally depend on the received optical power (ROP) to the photodetector. Figure 2(c) presents two cases where the ROP is sub-optimally chosen. If the ROP is too high, the threshold n' will be low and the ReLU function will be impacted by the non-linear IL slope below its -3 dB transmission point, as indicated through e_{IL}. Conversely, for a too low ROP, the threshold n'' will be too high and the neural data signal will find itself above the -3 dB point of the IL slope, leading to a non-linear response at the maximum of the IL transmission (m_{IL}). The ROP has to be optimized to minimize the error ξ in the ReLU function, which is given by the deviation from its linear slope according to

$$\xi = \int_{\nu_\lambda}^{\nu_\lambda+\Delta\nu} \left|T(\nu) - L(\nu)\right| d\nu \tag{5}$$

and is indicated by the red areas at e_{IL} and m_{IL} in Fig. 2(c). The condition for an ideal threshold n^* is sketched in Fig. 2(d). The processed neural data, which is represented by the photocurrent i^* of the optical interleaver output after frequency-to-intensity conversion k_{IL} at its linear slope, is correctly rectified while no additional distortion is incurred. This further requires a careful match of the frequency deviation $\Delta\nu$ to the free spectral range of the IL, to retain the neural signal s_v within the linear region $L(\nu)$. At the same time, $\Delta\nu$ cannot be made arbitrarily small to ensure linear operation since this would result in a degraded signal-to-noise ratio. A careful trade-off that balances non-linear distortion and excess noise needs to be made.

Although a photonic approach involving opto-electronics is required, we will show in Section V that a conversion between the optical and electrical domains between the layers of the ONN does not necessarily relate to complex or inefficient interfaces.

We characterized the response of this ReLU function experimentally. For this, a vector network analyzer (VNA) drives an optical inphase/quadrature modulator (IQM) through a wideband radio frequency (RF) 90° hybrid. This enables us to generate an optical single-sideband frequency sweep. A 25/50 GHz IL (Optoplex IL-C0SBFAS004) with a slope of k_{IL} = 5.43 dB/GHz is then inserted between IQM and PIN photoreceiver (HP 11982A) with a noise equivalent power of

This article has been accepted for publication in IEEE/OSA Journal of Lightwave Technology. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JLT.2024.3413976

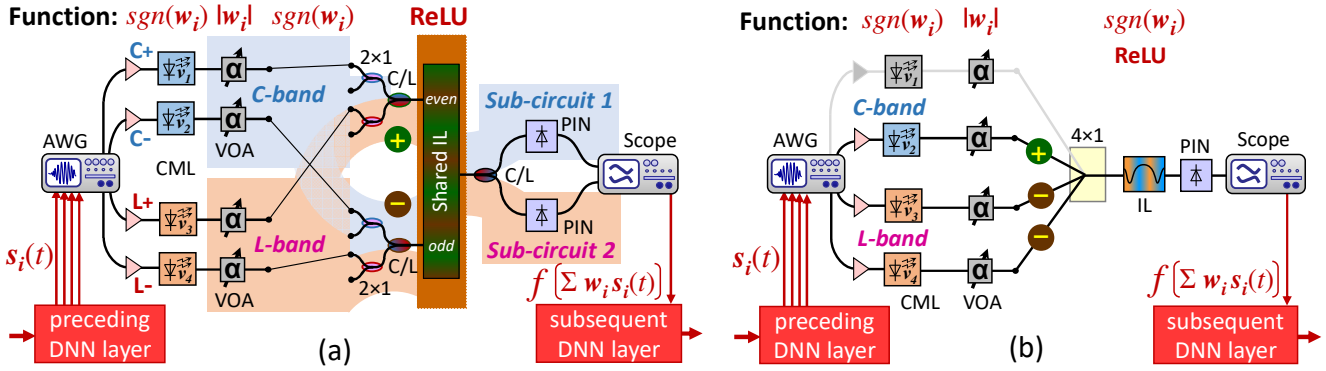

Fig. 9. Experimental setup to evaluate an ONN based on frequency-domain ReLU with (a) multiple neural sub-circuits using common ReLU hardware, and (b) simultaneous weighting and non-linear thresholding when scaling up the number of neural inputs.

30 pW/√Hz. The response of this link is then acquired through the VNA. Figure 3 presents the corresponding transfer function when spectrally allocating the optical source at the onset of an IL transmission window while adjusting the launch power with an attenuator to accomplish clipping of the filtered response. For an average ROP below -10.5 dB, we see a suppression towards lower frequencies $f < f_K$, with $f_K = 0$ GHz. Above f_K a linear slope is obtained, which together with the suppressed region resembles a ReLU function.

B. *Neural Network Model and Application Setting*

In our work, we use the Iris flower classification problem to evaluate the performance of the proposed photonic neuron. For this, a dataset of 150 flowers is distributed in three classes: Iris setosa, Iris versicolor, and Iris virginica [39]. Each flower sample has four features, consisting of information on its sepal (length and width) and petal (length and width). For that purpose, we use a multilayer perceptron (MLP) architecture (Fig. 1) with three hidden layers besides the input and output layers. In each hidden layer, there are four nodes, except layer 3 with only three nodes. As a benchmark, the DNN has been trained offline with 100 epochs using the Adam optimization [40]. There are in total 55 trainable parameters with a distribution of 20, 20, and 15 parameters for layer 1, 2 and 3, respectively. It shall be noted that the activation function does not occupy the parameters. Moreover, a common partition of 80% for training and 20% for testing has been implemented in the DNN. In addition, a batch size of 20 and a learning rate of 10^{-3} have been used in the training optimization.

Each input (x_i) of the four nodes in the first layer represents one feature of the flower. Each neuron in the NN, featuring the architecture shown in Fig. 1(b), is characterized by weighting coefficients (w_i) as an outcome of the training process. Afterwards, we perform a MAC operation and apply a bias (b) to their sum (Σ). Then, we utilize the nonlinear activation function (f) and the resulting output is sent to the subsequent NN layer. At the final output layer, a softmax function (z_i) is applied to determine the prediction probability. For performance evaluation, the outcome of the ONN involving the photonic ReLU activation function will be compared to the ideal DNN implementation.

III. Comparison of ReLU Functions for ONNs

First, we evaluate photonic ReLU functions based on two schemes (Fig. 4), including (*i*) the proposed frequency-domain ReLU with optical IL and (*ii*) a directly modulated laser (DML).

In the first implementation, the weighted sum Σ from the digital linear part of a neuron within the DNN is frequency-coded on an optical wavelength λ. The Σ_{FM} signal is then processed by the transfer function of the IL in combination with the subsequent photodetector, as previously discussed in Section II.A. In the second case of employing a DML, we utilize the electro-optic conversion characteristics to realize a ReLU function. This is done by adding a bias current on the weighted sum signal Σ_{RF}, to align the neural signal to the light-current (*L-I*) characteristics of the DML which resemble a ReLU function around the threshold current I_{th} of the DML.

Figure 5 introduces the experimental setup to evaluate these two ReLU functions. The weighted sum data $\Sigma w_i s_i(t)$ from node 0, layer 2 of the DNN (see Fig. 1 for reference) is programmed in an arbitrary waveform generator (AWG), without being digitally processed by an activation function. The 1 Gb/s neural data signal then drives a 10-GHz butterfly CML (Finisar DM80) operating at 1600.6 nm. The inset in Fig. 5 shows the optical output of the CML, which features a low intensity extinction ratio of 0.3 dB, as it is expected for an optically frequency modulated (FM) signal. The CML output is then processed by the optical 25/50 GHz IL to perform ReLU operation through frequency-selective demodulating of the FM signal to an intensity modulated (IM) signal, using the linear IL transmission slope around its quadrature point. Figure 6(a) depicts the optical spectra of CML before (α) and after passing the IL (β), along with the transfer function of the IL channel (γ). The output of the photonic ReLU is received by a PIN receiver, which is responsible for clipping the optical output of the IL at its lower intensity levels, and digitized through a real-time oscilloscope. The digitized signal is then fed back to the DNN implementation for evaluation of the error and NN accuracy.

For the second ReLU implementation, a 10-GHz butterfly DML (Youopto YDBK) operating at 1554 nm was used. According to its threshold current I_{th} of 6.5 mA, we

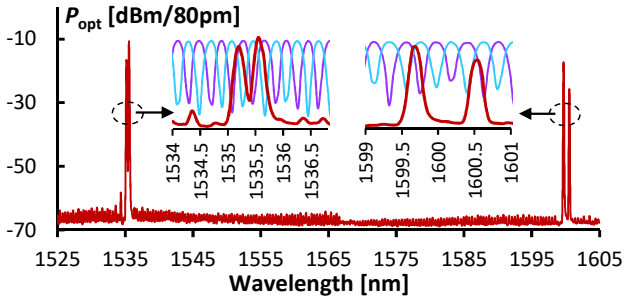

Fig. 10. Optical spectra of four CMLs in C- and L-band after shared IL hardware. The insets highlight the spectral allocation of the CMLs to the even and odd channels of the IL, realizing the weight polarity assignment.

investigated two different bias currents for ONN operation, as will be motivated shortly. The first bias resembles a sub-threshold current of 5.5 mA (δ), followed by a bias current of 11.2 mA (ζ) above I_{th}. The corresponding optical spectra are reported in Fig. 6(b) and indicate a wider spectral feature for the sub-threshold condition.

Figure 7(a) reports the normalized neural signal at the output of the ReLU unit based on CML+IL for all 150 Iris flower samples (●). The processed ReLU output of the hybrid ONN is in good agreement with the digitally computed output of the DNN (○). This remarkable match is evidenced by the very similar eye diagrams between the reference ReLU output of the DNN (Fig. 7(c)) and the CML+IL based ReLU unit (Fig. 7(d)).

In contrast, the ReLU operation based on the DML (▲) introduces a larger error when applying the low sub-threshold bias current to rectify the neural data signal, as depicted in Fig. 7(b). This error is explained by the gain switching effect under sub-threshold conditions, which are noticeable through the pulsed DML emission in the corresponding eye diagram (Fig. 7(e)) or the broadened optical spectrum (Fig. 6(b, δ)). When instead increasing the DML bias above I_{th} in an attempt to overcome this effect, the ReLU operates in the linear *L-I* region without gain switching artifacts. This, however, prevents us from performing rectification, as it becomes clear from the up-shifted eye diagram (Fig. 7(f)). Therefore, the DML is not suitable to perform the ReLU function.

Figure 8 summarizes the resulting error in the ReLU operation for both, CML+IL and DML based implementations. The error ε is defined as the absolute difference between the normalized optical output signals of the photonic ReLU and that of the DNN. The error under CML+IL operation (κ) is expectedly lower, with an average error of $\bar{\varepsilon} = 0.14$ ($3\sigma = 0.36$) when compared to the DML operation (ς), for which the error is distributed in a range of $0 < \varepsilon < 0.7$.

Additionally, we fed the output of the photonic ReLU back to the DNN layer, meaning that the digitized measurement data is propagated forwardly to the output layer of MLP architecture. For the chosen Iris flower classification problem, we achieve an accuracy of 90% and 62% by using the CML+IL and DML based ReLU implementations, respectively. This emphasizes that the CML+IL is suitable to perform the photonic ReLU function, leading to a 3% penalty

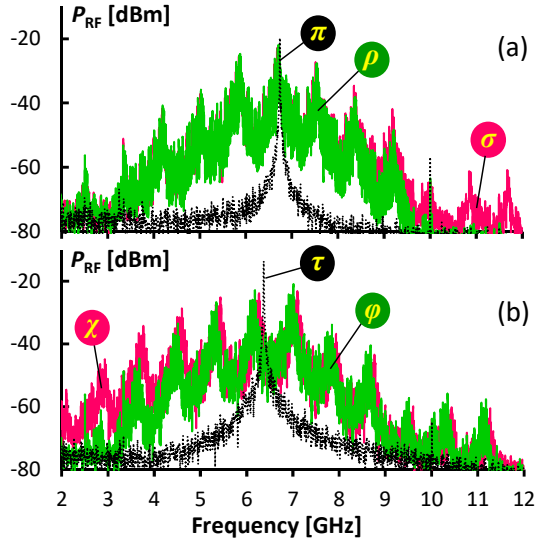

Fig. 11. Heterodyned neural spectra before (σ, χ) and after (ρ, φ) the shared IL for (a) negatively and (b) positively weighted neural signal, respectively.

in terms of accuracy when compared to the DNN, which itself features a 93% accuracy after offline training. Therefore, further investigations will build on the CML+IL combination as a photonic ReLU.

It shall be stressed that the NN training did not take component imperfections of the ONN layer into consideration, meaning that the DNN training on the neurons is directly transferred to the optical layer, regardless of the actual sub-system implementation for the involved photonic neuron.

IV. Adoption of the Frequency-Domain ReLU in ONNs

The CML+IL based ReLU scheme will be further evaluated in ONN environments with simultaneous weighting operation and under the aspect of adopting shared optical hardware among multiple neuromorphic circuits.

A. Collapsing multiple neural sub-circuits over a common optical-layer hardware

In the first scenario, for which the experimental setup is reported in Fig. 9(a), the IL is used as a shared ReLU element for two neural sub-circuits. All synaptic emitters will share the same IL to provide a simplified and cost-optimized ONN layout where multiple neural circuits can share the hardware within a NN layer.

We use four CMLs, two of which are in the C-band (1535.2 and 1535.55 nm) and two are allocated to the L-band (1599.65 and 1600.5 nm). Both C-band CMLs define the synaptic emission of two nodes in layer 2 (*E* and *F* in Fig. 1). The corresponding wideband PIN receiver yields the output of the neuron of node 2, layer 3 (*H*). Similarly, both L-band CMLs represent the same emitters (*E* and *F*) and the processed result is dedicated to the neuron at node 0, layer 3 (*G*). The optical spectrum of all CMLs after the shared IL is shown in Fig. 10 with the IL transmission function of even and odd channels reported as insets. The spectral allocation of the CMLs to the IL slopes again follows the synaptic weight assignment.

We send the data from DNN $s_i(t)$ of the corresponding neurons in layer 2 to drive the CMLs at 1 Gb/s. Assigning the

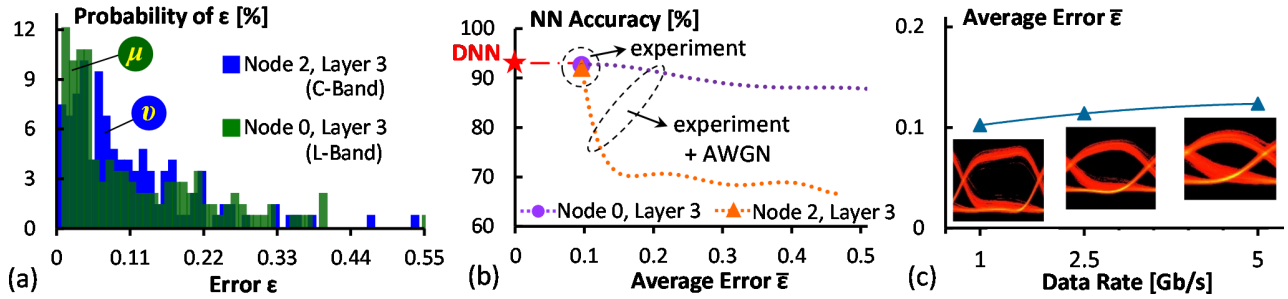

Fig. 12. (a) Error histogram for the optical NN output of node 0 (μ) and node 2 (υ) in layer 3, after performing optically weighted sum and ReLU function, and (b) corresponding NN accuracy for Iris flower classification at 1 Gb/s for an average error $\bar{\varepsilon}$ of 0.1 and the drop in accuracy through additive white Gaussian noise (AWGN) that has been artificially added to degrade $\bar{\varepsilon}$. (c) Average error $\bar{\varepsilon}$ as a function of the data rate.

sign of the weight, $sgn(w_i)$, is accomplished through an alignment of the frequency-coded signals to the respective positive or negative IL slopes, taking advantage of its periodic transfer function. Moreover, a frequency-agnostic variable optical attenuator (VOA) is placed in each CML output to set the weight magnitude $|w_i|$. The two neurons (E and F) have alternating signs (w_+,w_-) based on the DNN training. In this second scenario, we used 1×2 splitters to enable the required weight combinations and connections towards the even (+) and odd (–) input of the shared IL, while C/L waveband splitters are separating the two ONN sub-circuits. Finally, the signal is detected by PIN receivers in either waveband and the output is digitized through a real-time oscilloscope to determine the accuracy of the ONN.

Figure 11 shows how the synaptic weight assignment in correlation with the frequency-coded ReLU operation is conducted by using a CML with a chirp parameter of 2.9. The spectra summarize the continuous-wave beat notes (π, τ) for the unmodulated C-band CML and a reference laser, resulting in a down-conversion frequency of ~6.6 GHz. In addition, the RF spectra report the heterodyned neural signals before (σ, χ) and after (ρ, φ) the shared IL for a negative (Fig. 11(a)) and a positive (Fig. 11(b)) sign setting. For a proper rectification in case of a negative sign, the higher frequencies of the frequency-coded neural signal are suppressed by the falling slope of the IL (ρ), as shown in Fig. 11(a). Conversely, the rising slope of the IL suppresses the lower frequencies of the neural signal (φ in Fig. 11(b)) for the positive sign setting.

Figure 12(a) reports the error histogram for the two ReLU sub-circuits with shared IL. The result for both neurons in layer 3, node 0 (μ, L-band) and node 2 (υ, C-band), has the same average error $\bar{\varepsilon}$ as low as 0.1. We further calculated the accuracy of Iris flower prediction by propagating the optical measurement data through the remaining portion of the DNN. Figure 12(b) summarizes the accuracy as a function of $\bar{\varepsilon}$, taking input from the experiment and from an evaluation where we added additional noise to the experimental data at the digital domain in order to simulate further degradation of the neural signal. For both neurons, we obtain an accuracy of 92% for the experiment at 1 Gb/s, without modifying the NN training procedure to take the specifics of the involved optical elements into consideration. This result stands very close to the accuracy of 93% for the DNN. When $\bar{\varepsilon}$ increases due to an artificial noise degradation, we see a rather large drop in

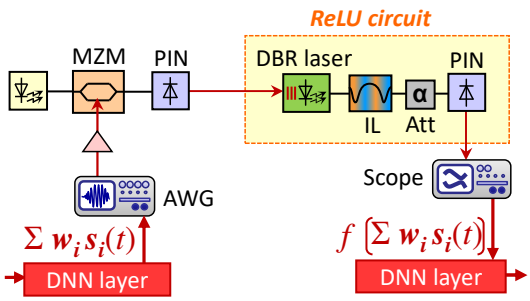

Fig. 13. Concept and experimental setup for an o/e/o conversion between the linear and nonlinear part of a photonic neuron, employing a DBR laser directly driven by a PIN photodiode current. The same photonic ReLU scheme is used to evaluate the direct-drive scheme.

accuracy for node 2 (▲). On the other hand, the accuracy for node 0 (●) depletes just slightly. This result can be understood in view of the distinct noise tolerance of the involved neurons. It would require precise optical implementations for the particularly noise-sensitive neurons, while other sub-circuits could feature more relaxed conditions.

Figure 12(c) shows the relationship between $\bar{\varepsilon}$ and the neural data rates for the noise-sensitive node 2 at layer 3. We see a slight increase in $\bar{\varepsilon}$ from 0.10 to 0.12 for a range of 1 to 5 Gb/s in data rate where the accuracy drops to 90% at 5Gb/s.

B. *Scaling simultaneous weighting and ReLU operation*

In the second scenario, we investigate whether the combination of ReLU function and weight setting can scale. The corresponding experimental setup is presented in Fig. 9(b) and features a single neural circuit with three synaptic emitters. We weight the outputs of the neurons in layer 1 (A, B, D in Fig. 1) and sum them, resulting in the output of node 0, layer 2 (E) after performing the ReLU function. We use two CMLs in the L-Band to represent the emitters of the two nodes in layer 1 (A [w_1] and D [w_4]) and one CML in the C-band for the third node (B [w_2]). In the same manner as the first scheme, the $s_i(t)$ data is sent to the CMLs of the respective neurons at 1 Gb/s. The two neural data streams in the L-band have negative signs (w_-), while the signal in the C-band has positive polarity (w_+). The three synaptic signals are combined with a colourless 4×1 splitter and sent to the IL to perform the ReLU activation function. Finally, the signal is detected by a PIN receiver and we again determine the accuracy of the ONN after digitization of the neural signal.

In this scenario, we achieve an average error $\bar{\varepsilon}$ of 0.2. This

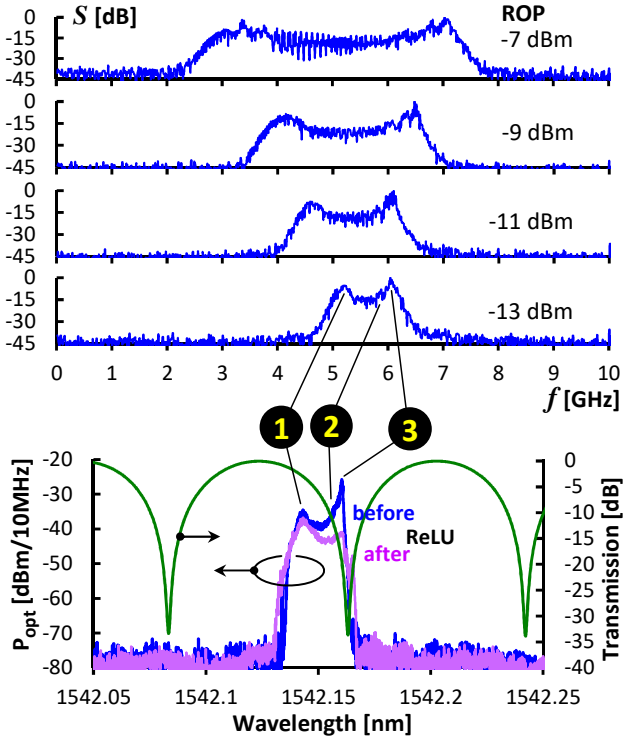

Fig. 14. Spectrogram of the DBR laser output and spectral alignment of the DBR laser to the IL transfer function to realize a photonic ReLU function.

is higher than what we obtained in the first scenario. When we feed the optical measurement signal back to the MLP architecture of the DNN, we still achieve an accuracy of 91% for the Iris flower classification challenge. However, this might be the result of a noise-insensitive neuron configuration, given the rather high $\bar{\varepsilon}$ value. The latter is explained by the fact that the integration of the weight polarity setting with the ReLU function does not resemble the original function of the neuron since the ReLU is not applied to the weighted sum. Precise neural signal processing in a scaled-up ONN would therefore require a clear split between the linear and the nonlinear sub-circuits within the photonic neuron. An efficient method to do so will be introduced in the next Section.

V. Translucent Concatenation of Weighted Sum and Nonlinear Activation

When realizing the nonlinear activation function of the photonic neuron as a dedicated photonic circuit, the necessity for signal translation between the electrical and optical layers at the boundary of the linear and nonlinear neural sub-circuits arises. This optic/electric/optic (o/e/o) translation can quickly become a burden when scaling up the ONN. The following experimental demonstration aims to prove that a detriment in energy consumption, such as it arises from signal conditioning during o/e/o conversion, can be largely avoided.

This o/e/o process shall be accomplished by a direct-drive scheme where a highly efficient FM transmitter is directly driven by the photocurrent of a preceding photodiode (Fig. 13). Specifically, a tandem of PIN photodiode and distributed Bragg reflector (DBR) laser constitute a translucent intensity-

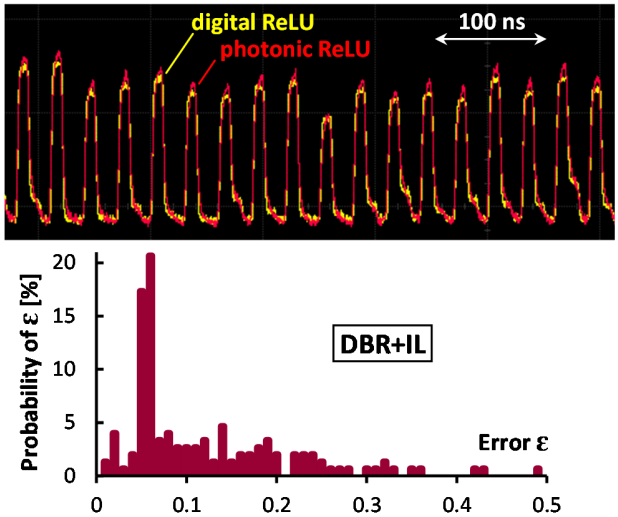

Fig. 15. ReLU output from the optical and all-digital NN implementation, and corresponding error histogram for the photonically implemented ReLU function, resulting to a low average error $\bar{\varepsilon}$ of 0.1.

to-frequency converter where the detected photocurrent is translated to an optical frequency-coded representation according to Eq. (1), with $s(t)$ now being the photocurrent that represents the optically computed weighted sum.

Figure 14 characterizes the efficiency of this frequency translation process through the DBR laser (Youopto YTET) for a neural input with a 100 Mbaud information rate. For this purpose, the DBR laser output has been directly acquired with a high-resolution optical spectrum analyzer and has been additionally heterodyned with a reference laser to obtain its FM spectrum at a down-conversion frequency of ~5.5 GHz. As Fig. 14 proves, a ROP of -7 dBm to the PIN photodiode that drives the DBR section leads to a frequency-translated neural signal with a deviation of $\Delta\nu = 4$ GHz at the DBR laser output, meaning a ~50-fold spectral expansion from the amplitude to the frequency space – without necessitating an electronic amplifier. This remarkable ratio between neural symbol rate and frequency deviation is propelled by the excellent spectral tuning efficiency of DBR lasers. Given the passive and athermal nature of the IL and the direct-drive capability of the DBR laser, which requires no means of RF amplification, the energy efficiency of the entire ReLU function is determined mainly by the optical gain for the laser source and the post-amplification stage of the photodetector. Towards these, the currently employed DBR laser was supplied at a bias of 40 mA at a forward voltage of 1.2V for its gain section, while the TIA of the photodetector was supplied at 25 mA at a 3V rail. The static power consumption of the ReLU is therefore ~125 mW, equivalent to 1.3 nJ/operation given the currently low information rate of 100 Mbaud. This efficiency value would decrease significantly once DBR lasers with higher electro-optical modulation bandwidth, such as [41], become widely available.

The spectrum in Fig. 14 further proves that the FM signal at the DBR laser output, overlaid with the transmission function of the IL of the ReLU circuit, features a deviation wide enough to obtain the desired rectification effect, which in case

of the spectrum shown in Fig. 14 relates to a (partial) suppression of the levels 2 and 3 of the neural data (see Fig. 7(c) as a reference) after passing the photonic ReLU function.

We have then evaluated the error resulting from the insertion of this o/e/o conversion. As Fig. 13 details, the weighted sum signal $\Sigma w_i s_i(t)$ of node 0 layer 2 in the DNN (*E* in Fig. 1) is converted to an optical signal at 1550 nm by a Mach-Zehnder Modulator (MZM). This optical representation of the sum feeds a PIN photodetector, which then drives the DBR section of the laser that optically supplies the ReLU circuit. The DBR laser is aligned to the periodic IL transmission of the ReLU function through a fine-tuning of the bias current of the gain section of the DBR laser. After passing the ReLU, the signal is detected by a PIN photodiode and evaluated following the same methodology presented earlier.

A comparison between a photonically performed ReLU (red trace) and a digitally performed ReLU (yellow) is reported in Fig. 15. We see a good agreement between the optical and digital computation, apart from a bandwidth limitation in the optical signal. This is attributed to the limited electro-optic 3dB bandwidth of the DBR section, which was 270 MHz. However, the neural signal levels agree with the digital levels at the optimum decision sampling point. This is confirmed by the error histogram: By using the proposed direct-drive DBR scheme of the FM transmitter, we obtain an average error $\bar{\varepsilon}$ of 0.1. This is slightly lower than for the CML+IL based scheme evaluated earlier in Section IV. When feeding back the signal acquired at the ReLU output to the MLP architecture in the digital domain and forwardly propagating it to the output layer, we achieve an accuracy of 92% for Iris flower classification. This 1% penalty in comparison to the DNN proves that an efficient o/e/o conversion between two photonic signal processing circuits can be realized, without the need for electrical signal conditioning.

VI. Conclusion

We have experimentally demonstrated a photonic ReLU function in the frequency domain, leveraging optical frequency modulation in combination with chirped light sources or frequency-tunable lasers in combination with periodic optical filters, followed by a PIN photodiode. Based on our evaluation, the electro-optic DML approach is incompatible to perform the desired ReLU function. The proposed ReLU function centric to the CML+IL tandem has been integrated in a hybrid digital/optical NN implementation, showing the beneficial sharing of the ReLU hardware by multiple neural circuits. We have further proven that a penalty as low as 1% can be achieved in terms of accuracy when solving the Iris flower classification problem. Moreover, we have evaluated the simultaneous weight setting through the same hardware used to perform the ReLU function. Though moderate errors have been achieved for a low number of neural inputs, a scalable solution calls for separate sub-circuits dedicated to linear and nonlinear processing. Towards this direction, we have experimentally demonstrated a simplified and efficient translucent o/e/o interface between weighted sum computation and ReLU activation, essentially bonding these two functions with a tandem of a PIN photodiode and DBR laser, with the latter being directly driven through the photocurrent of the preceding photodiode. Although larger bandwidths are required for DBR modulation, it promises an energy-efficient interface with small footprint due to its amplifier-less nature. The investigation of more compact optical elements replacing the currently used optical interleaver is left for future work.

References

[1] N. Jones, "The information factories," *Nature*, vol. 561, pp. 163-166, Sep. 2018.

[2] P. Messina, "The exascale computing project," *Computing in Science & Engineering*, vol. 19, no.3, pp. 63-67, May-Jun. 2017.

[3] J. von Neumann, "First draft of a report on the EDVAC," *IEEE Ann. Hist. Comput.*, vol. 15, no. 4, pp. 27-75, Oct. 1993.

[4] E. Track, N. Forbes, and G. Strawn, "The end of Moore's law," *IEEE Comput. Sci. Eng.*, vol. 19, no. 2, pp. 4-6, Mar.-Apr. 2017.

[5] B. J. Shastri, A. N. Tait, T. F. de Lima, M. A. Nahmias, H.-T. Peng, and P. R. Prucnal, "Principles of neuromorphic photonics," in *Unconventional Computing*. New York, NY, USA: Springer, 2018, pp. 83–118.

[6] Y. Shen *et al.*, "Silicon photonics for extreme scale systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 245-259, Jan. 2019.

[7] J. D. Kendall and S. Kumar, "The building blocks of a brain-inspired computer," *Appl. Phys. Rev.,* vol. 7, no. 1, Jan. 2020, Art. no. 011305.

[8] J. Hasler and B. Marr, "Finding a roadmap to achieve large neuromorphic hardware systems," *Front. Neurosci.*, vol. 7, no. 118, pp. 1-29, Sep. 2013.

[9] C. D. Schuman, S.R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, "Opportunities for neuromorphic computing algorithms and applications," *Nat. Comput. Sci.*, vol. 2, pp. 10-19, Jan. 2022.

[10] T. Baba *et al.*, "Proposal of disruptive computing (A computing-domain-oriented approach)," *Jpn. J. Appl. Phys.*, vol. 59, Apr. 2020, Art. no 050503.

[11] X. Guo, J. Xiang, Y. Zhang, and Y. Su, "Integrated neuromorphic photonics: Synapses neurons and neural networks," *Adv. Photon. Res.*, vol. 2, no. 6, Jun. 2021, Art. No. 2000212.

[12] E. Goi, Q. Zhang, X. Chen, H. Luan, and M. Gu, "Perspective on photonic memristive neuromorphic computing," *PhotoniX*, vol. 1, no. 3, Mar. 2020.

[13] M. A. Nahmias, B. J. Shastri, A. N. Tait, T. F. de Lima, and P. R. Prucnal, "Neuromorphic photonics," *Opt. Photon. News*, vol. 29, no. 1, pp. 34-41, Jan. 2018.

[14] T. F. de Lima *et al.*, "Machine learning with neuromorphic photonics," *J. Lightw. Technol.*, vol. 37, no. 5, pp. 1515-1534, Mar. 2019.

[15] C. Huang *et al.*, "Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems," in *Proc. Opt. Fib. Comm. Conf. (OFC)*, San Diego, United States, Mar. 2020, paper Th4C.6.

[16] G. Giamougiannis *et al.*, "Neuromorphic silicon photonics with 50 GHz tiled matrix multiplication for deep-learning applications," *Adv. Photon.*, vol. 5, no. 1, Feb. 2023, Art. no. 016004.

[17] W. Zhang *et al.*, "Broadband physical layer cognitive radio with an integrated photonic processor for blind source separation," *Nat. Commun.*, vol. 14, Feb. 2023, Art. no. 1107.

[18] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nat. Photon.*, vol. 11, pp. 441-446, Jun. 2017.

[19] A. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 3427-3439, Nov. 2014.

[20] B. Shi, N. Calabretta, and R. Stabile, "Deep neural network through an InP SOA-based photonic integrated cross-connect," *J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan.-Feb. 2020, Art no. 7701111.

[21] M. V. Stephanie *et al.*, "SOA-REAM assisted synaptic receptor for weighted-sum detection of multiple inputs," *J. Lightw. Technol.*, vol. 41, no. 4, pp. 1258-1264, Feb. 2023.

[22] A. Nguyen, K. Pham, D. Ngo, T. Ngo and L. Pham, "An analysis of state-of-the-art activation functions for supervised deep neural network,"

in *Proc. Int. Conf. System Science and Engineering (ICSSE)*, Ho Chi Minh City, Vietnam, Sep. 2021, pp. 215-220.

[23] J. K. George *et al.*, "Neuromorphic photonics with electro-absorption modulators," *Opt. Express*, vol. 27, no. 4, pp. 5181-5191, Feb. 2019.

[24] M. M. P. Fard *et al.*, "Experimental realization of arbitrary activation functions for optical neural networks," *Opt. Expr.*, vol. 28, no. 8, pp. 12138–12148, Apr. 2020.

[25] I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, Jan.-Feb 2020, Art. no. 7700412.

[26] J. R. Rausell Campo and D. Pérez-López, "Reconfigurable activation functions in integrated optical neural networks," *J. Sel. Topic Quantum Electron.*, vol. 28, no. 4, Jul.-Aug. 2022, Art. no. 8300513.

[27] Y. Huang, W. Wang, L. Qiao, X. Hu, and T. Chu, "Programmable low-threshold optical nonlinear activation functions for photonic neural networks," *Opt. Lett.*, vol. 47, no. 7, pp. 1810-1813, Apr. 2022.

[28] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, "An all-optical neuron with sigmoid activation function," *Opt. Express*, vol. 27, no. 7, pp. 9620-9630, Apr. 2019.

[29] B. Shi, N. Calabretta, and R. Stabile, "InP photonic integrated multi-layer neural networks: Architecture and performance analysis," *APL Photonics,* vol. 7, Jan. 2022, Art. no. 010801.

[30] Q. Li *et al.*, "SOA-based all-optical neuron with reconfigurable nonlinear activation functions," in *Proc. Conference on Lasers and Electro-Optics (CLEO),* San Jose, CA, USA, May 2022, paper SF4F.6.

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

[32] Z. Fu, Z. Wang, P. Bienstman, R. Jiang, J. Wang, and C. Wu, "Programmable low-power consumption all-optical nonlinear activation functions using a micro-ring resonator with phase-change materials," *Opt. Expr.*, vol. 30, no. 25, pp. 44943-44953, Dec. 2022.

[33] 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," *Opt. Mater. Expr.,* vol. 12, no. 3, pp. 970-980, Mar. 2022.

[34] N. Takahashi, *et al.*, "Optical ReLU using membrane lasers for an all-optical neural network," *Opt. Lett.*, vol. 47, no. 21, pp. 5715-5718, Nov. 2022.

[35] J. Crnjanski, M. Krstić, A. Totović, N. Pleros, and D. Gvozdić, "Adaptive sigmoid-like and PReLU activation functions for all-optical perceptron," *Opt. Lett.*, vol. 46, no. 9, pp. 2003-2006, May 2021.

[36] G. H. Y. Li *et al.*, "All-optical ultrafast ReLU function for energy-efficient nanophotonic deep learning," *Nanophotonics*, vol. 12, no. 5, pp. 847-855, May 2022.

[37] M. V. Stephanie, L. Pham, A. Schindler, M. Waltl, T. Grasser and B. Schrenk, "All-Optical ReLU as a Photonic Neural Activation Function," in *Proc. IEEE Photonics Society Summer Topicals Meeting Series (SUM)*, Sicily, Italy, Jul. 2023, paper MF4.4.

[38] M. V. Stephanie, L. Pham, A. Schindler, M. Waltl, T. Grasser and B. Schrenk, "Neural Network with Optical Frequency-Coded ReLU," in *Proc. Opt. Fib. Comm. Conf. (OFC)*, San Diego, United States, Mar. 2024, paper M4C.2.

[39] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179-188, Sep. 1936.

[40] O. Konur, D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations*, San Diego, United States, May. 2015, pp. 1-15.

[41] M. Pantouvaki, C.C. Renaud, P. Cannard, M.J. Robertson, R. Gwilliam, and A.J. Seeds, "Fast Tuneable InGaAsP DBR Laser Using Quantum-Confined Stark-Effect-Induced Refractive Index Change," *J. Sel. Topics in Quantum Electron.*, vol. 13, no. 5, pp. 1112-1121, Sep. 2007.