A Photonic Recurrent Neuron for Time-Series Classification

George Mourgias-Alexandris, Nikolaos Passalis, George Dabos, Angelina Totović, Anastasios Tefas
and Nikos Pleros

Abstract— **Neuromorphic photonics has turned into a key research area for enabling neuromorphic computing at much higher data-rates compared to their electronic counterparts, improving significantly the (Multiply-and-Accumulate) MAC/sec. At the same time, time-series classification problems comprise a large class of Artificial Intelligence (AI) applications where speed and latency can have a decisive role in their hardware deployment roadmap, highlighting the need for ultra-fast hardware implementations of simplified Recurrent Neural Networks (RNN) that can be extended in more advanced Long-Short-Term-Memory (LSTM) and Gated Recurrent Unit (GRU) machines. Herein, we experimentally demonstrate a novel Photonic Recurrent Neuron (PRN) for classifying successfully a time-series vector with 100-psec optical pulses and up to 10Gb/s data speeds, reporting on the fastest all-optical real-time classifier. Experimental classification of 3-bit optical binary data streams is presented, revealing an average accuracy of >91% and confirming the potential of PRNs to boost speed and latency performance in time-series AI applications.**

Index Terms— **neuromorphic photonics, neuromorphic computing, optical neural networks, recurrent neural networks, programable photonics.**

I. INTRODUCTION

NEUROMORPHIC photonics has experienced a recent surge of interest during the last few years, raising expectations for neuromorphic hardware that can perform directly in the optical domain at orders of magnitude higher data rates compared to the corresponding electronic solutions [1].This can yield higher MAC/sec computational speeds that can subsequently translate into important energy and area efficiency gains, taking also advantage of the rapid progress experienced in photonic integration [2], [3]. This reality has stimulated the deployment and demonstration of a whole new class of optical modules for use as basic building blocks in photonic artificial neurons, including optical weighting units [4]–[9] and non-linear activation functions [10]–[14], leading also to innovative neuromorphic photonic architectures [15]–[23] and training processes [24]–[27]. The main emphasis in this effort has so far relied on feed-forward [15], convolutional [22] and spiking [17] neural network layouts, with experimental demonstrations typically restricted to proof-of-principle validations that are subsequently utilized in true application scenarios by means of ex-situ benchmarking procedures [9], [15], [28], where offline post-processing of the data is required. RNNs, which form typically the cornerstone of all necessary Deep Learning (DL)-based time-series analysis applications are still attempting their very first elementary steps in migrating to their optical version. To this end, a large set of speed- and latency-critical time-series applications like pattern classification, forecasting, text processing, natural language processing, finance applications and routing address identification, have only been partially addressed in the optical landscape, mainly through photonic Reservoir Computing (RC) schemes [18]–[21]. However, RC circuits are well-known as a special class of recurrent networks that is not used in the AI field and can't migrate to advanced LSTM and GRU machines. On top of that, time-series analysis and classification through photonic RC has been so far addressed only via ex-situ post-processing of the data, where the weighting stage is realized in the electrical domain. As a result, RC classifiers cannot translate the low-latency and high-speed credentials of optics into tangible real-time advantages in the AI application segment. Furthermore, the speed and latency performance for this type of problems in electronic layouts is defined by the overall speed and latency values supported by the underlying electronic hardware employed. As such, the use of a Tensor Processing Unit (TPU) or Graphic Processing Unit (GPU) results in speeds that range between a few hundreds of MHz up to a few GHz [2], [3], while the use of analogue neuromorphic layouts allows for much lower speed credentials that can hardly exceed a few tens of MHz [29]–[31]. In both cases, latency goes into the msec regime.

In this work, we experimentally demonstrate for the first-time, to the best of our knowledge, an all-optical PRN that provides successful time-series classification directly in the optical domain and at speeds up to 10Gb/s, without requiring any post-processing since the weighing is realized in the optical domain as depicted in Fig. 1(a). The proposed all-optical PRN comprises a novel space rotator module followed by a neural classifier formed by a sigmoid activation element incorporated

G. Mourgias-Alexandris, N. Passalis, G. Dabos, A. Totović, A. Tefas and N. Pleros are with the Department of Informatics, Aristotle University of Thessaloniki, 54621 Thessaloniki, Greece (e-mail: mourgias@csd.auth.gr, passalis@csd.auth.gr, ntamposg@csd.auth.gr, angelina@auth.gr, tefas@csd.auth.gr, npleros@csd.auth.gr).

into a recurrent feedback loop. The role of the space rotator is to transform the incoming optical time-series vector in order to reside within the first quadrant of the corresponding coordinate system, where the optical neural classifier can be successfully applied for classifying the input sequence. The employed space rotator benefits from a trainable rotation mask that is kept constant during inference. Successful time-series classification has been experimentally presented for a 3-bit data sequence with 100psec-long optical pulses and for both a Return-to-Zero (RZ) modulation scheme at 3.3Gb/s as well as a Non-Return-to-Zero (NRZ) format at 10Gb/s, revealing an average accuracy of >90%.

II. CODING SCHEME & PRN ARCHITECTURE

The layout of typical all-optical recurrent neuron is illustrated in Fig. 1(a). The N inputs $x_1...x_n$ are weighted in the optical domain by the means of optical attenuators that apply the corresponding weighting value $w_1...w_n$. Then, weighted summation of input signals is forwarded into the photonic Activation Unit (AU). One copy of the output signal is weighted by the weighting value w_r and summed with the input signals, realizing in this way the recurrent loop of the photonic neuron. The second copy of the optical signal forms the output of the photonic neuron y_{out}. A special class of RNNs known as RC is depicted in Fig. 1(b) The N inputs $x_1...x_n$ are multiplied by a periodic mask signal with a period τ before entering the Non-Linear (NL) element to realize the θ-spaced N virtual nodes. Then, the optical signal of each node is converted to electrical domain to perform the network training and apply the corresponding weights. The summation of these electrical signals forms the output of the RC y_{out}, with every experimental demonstration of RC being relied on offline training [18]–[21].

RC circuits can be classified as having lower hardware complexity compared to RNNs, with their main advantage being the virtual nodes as well as the arbitrary weight configuration within the reservoir. However, RC circuitry is facing severe difficulties in scaling to higher complexity neural networks of increased functionality, like LSTM and GRU models, and this forms also the main reason for not having been still established in widely adopted, powerful DL architectures. On the other hand, RNN layouts allow for weighting and activation schemes similar to state-of-the-art DL networks, facilitating the adoption of powerful DL training techniques and
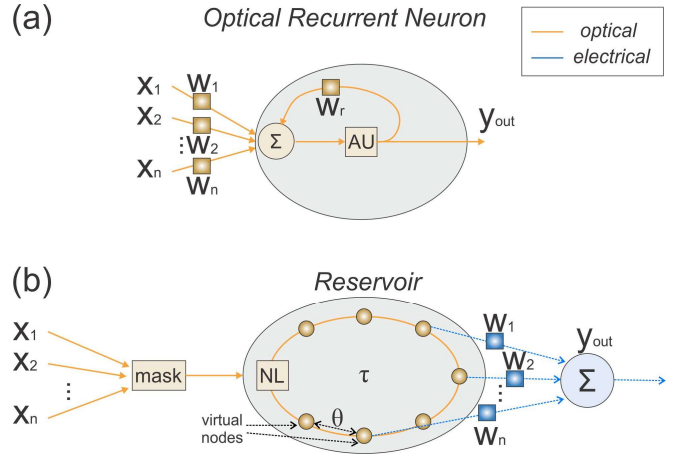

Fig. 1. Layout of (a) an all-optical recurrent neuron comprising n photonic weights, a summation stage followed by an Activation Unit (AU) as well as a recurrent loop with the corresponding weight, and (b) a photonic RC comprising an input mask, a Non-Linear (NL) element, virtual nodes and an electronic weighing and summation scheme.

their migration to more complex neural networks. Their compatibility with DL frameworks enables the acceleration of state-of-the-art applications, suggesting that the realization of all-optical RNN can pave the way towards all-optical LSTM and GRU modules while ensuring interoperability with well-known training models in a broad and diverse range of applications. Migrating from a single optical RNN into an optical LSTM or GRU requires of course additional sigmoid or tanh neural layers together with some gating mechanism to be incorporated, which have been already reported as constituent building blocks to operate in the all-optical domain [10]. Taking also into account that noise-resilient training methods are already extensively and successfully researched for all-optical RNNs with DL-compatible activations [24],[25], the introduction of all-optical RNNs into realistic AI applications like time series classification is expected to strengthen the credentials of optics towards penetrating the rich library of LSTM and GRU models.

The layout of the proposed PRN is illustrated in Fig. 2(a). The PRN comprises a space rotator and a photonic recurrent neuron, which in turn includes a weighting bank and the optical sigmoid activation unit reported in [10] incorporated into a fiber loop in order to enable the recurrent capabilities of the deployed neuron. The time-series x_t and the rotation mask m_t are both forwarded as input signals into the space rotator unit, providing at its output a rotated version $\tilde{x}_t$ of the input time-series x_t that
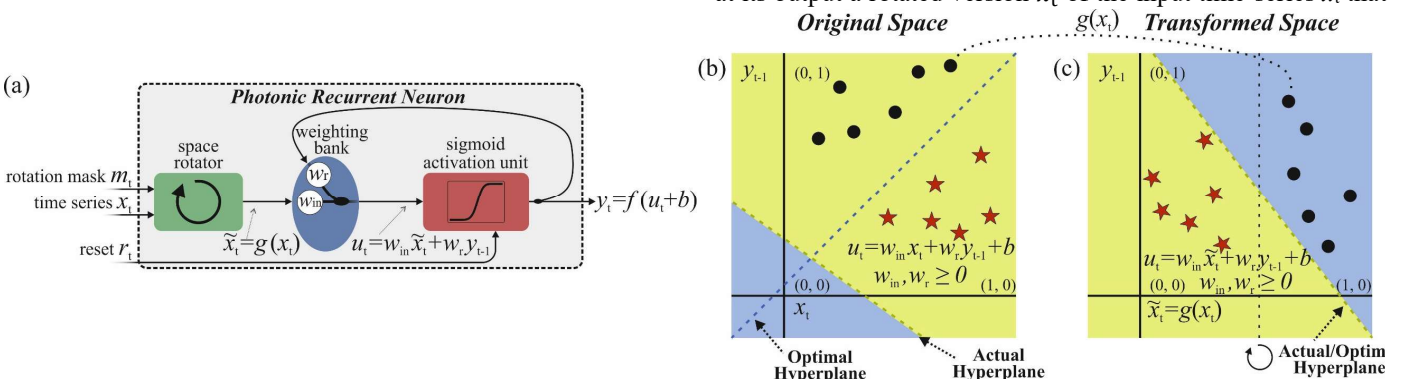

Fig. 2. (a) Layout of the proposed PRN, (b) the original sample space that supports hyperplanes only with positive parameters and (c) the transformed sample space, where the input sample space is rotated.

equals:

$$\tilde{x}_t = g(x_t) = 1 - x_t \qquad (1)$$

with g denoting the transfer function of the space rotator.

The rotated time-series $\tilde{x}_t$ gets then weighted by the input weight value w_{in} and is then multiplexed with the weighted recurrent signal $w_r y_{t-1}$ that equals zero at $t=0$, with y_{t-1} comprising the recurrent signal and w_r the weight of the recurrent signal. The summed output of the weighting bank $u_t = \tilde{x}_t w_{in} + y_{t-1} w_r$ is then injected into the sigmoid activation unit, where a bias signal b is additionally applied so that the output signal of the activation unit equals $y_t = f(u_t + b)$, with f denoting the transfer function of the sigmoid activation unit [10]. This signal is then split into 2 identical signals with the first copy providing the output of the PRN and the second signal being injected into a feedback loop so as to get weighted by w_r and delayed by $\Delta t=1$ in order to realize y_{t-1}. The weighted recurrent signal $w_r y_{t-1}$ is then multiplexed with the next time instance of the weighted time-series $w_{in}\tilde{x}_t$ prior reentering the activation unit, so that finally the output of the proposed PRN is governed by the following equation:

$$y_t = f(\tilde{x}_t w_{in} + y_{t-1} w_r + b) \qquad (2)$$

with w_{in} and w_r values ranging both within [0,1]. Before launching a new time-series into the PRN, the recurrent feedback from the previous signal will be discarded by resetting the PRN using the reset signal r_t.

The space rotator module is responsible for rotating the time-series input vector so as to reside on the first quadrant of the corresponding coordinate system, where the subsequent optically implemented neural classifier can be successfully applied for classifying the input sequence. Its role comes to safeguard the successful operation of an all-optical neuron where only positive weight and input signal values can be applied, which is typically encountered when all-optical nonlinear elements are used as high-speed activation modules. As such, the absence of the space rotator module when only positive weight values w_r and w_{in} are used would constrain the number of patterns that can be classified to the ones that can be separated using hyperplanes with positive parameters. To better understand this, consider a two-dimensional case, shown in Fig. 2(b). If only positive weights are used, then only negative slope decision boundaries can be implemented by the deployed neuron, the corresponding decision boundary is described by:

$$y_{t-1} = \left(-\frac{w_{in}}{w_r}\right)x_t + \left(-\frac{b}{w_r}\right) \qquad (3)$$

To overcome this hurdle, we propose the rotation function $g(x_t)$ that rotates the sample space by transforming the input data. Without loss of generality, we assume that $0 \leq x_t \leq 1$, which can be achieved simply by normalizing the input data, as frequently done in most AI applications, and that $0 \leq y_t \leq 1 + \epsilon$, which is also satisfied by definition, since a sigmoid activation function is used (ϵ refers to a small positive value). If the data can be separated using positive weights, then the deployed neuron could be used without any transformation, otherwise the input must be appropriately transformed. Let us consider two possible cases. First, assume that $w_{in} < 0$. Then, note that the neuron's response can be calculated as $u_t=\tilde{x}_t w_{in}+w_r y_{t-1}$. By also shifting the data $-x_t$ back to the $0...1$ interval, i.e., setting up the correct rotation axis, the following response function is obtained:

$$u_t = \tilde{x}_t(-w_{in}) + y_{t-1}w_r + (b + w_{in}) \qquad (4)$$

and all the weights are now positive (recall that $w_{in} < 0$). The case when $w_r < 0$ can be similarity handled by observing that in (3) only one of the w_{in} and w_r needs to be negative. Thus, the data can be always appropriately rotated to ensure that only positive weights will be required to separate them. Note that having a positive bias is not crucial, since the data can be always separated, even when $b = 0$, by setting the appropriate output recognition threshold. The proposed feature space rotation approach is depicted in Fig. 2(c). Note that after rotating the data, the decision surface can separate the data without using any negative weights. Moreover, the proposed method requires no changes to the training process, since it can be directly applied after training to the neuron, just by changing the sign of the negative weights, applying the transformation $g(x_t)$ to the corresponding inputs and setting then the new bias, as described by (4). The proposed rotation scheme can be also dynamically adapted, i.e., change through time by changing the rotation mask m_t without requiring to retrain the PRN.

III. ALL-OPTICAL 3-BIT TIME-SERIES CLASSIFICATION EXPERIMENT

A. Training of the PRN

The experimental validation of the all-optical time-series classification using the proposed photonic neuron has been carried out for three-symbol binary time sequences, where both the input and mask vectors comprise binary digits "0" and "1" with a pattern length of three bits. The PRN including the weighting bank and the fiber-loop-incorporated activation unit was trained to identify and classify the pattern "111". The proposed PRN was trained and evaluated for 3-bit pattern recognition problems using a dataset that was created after
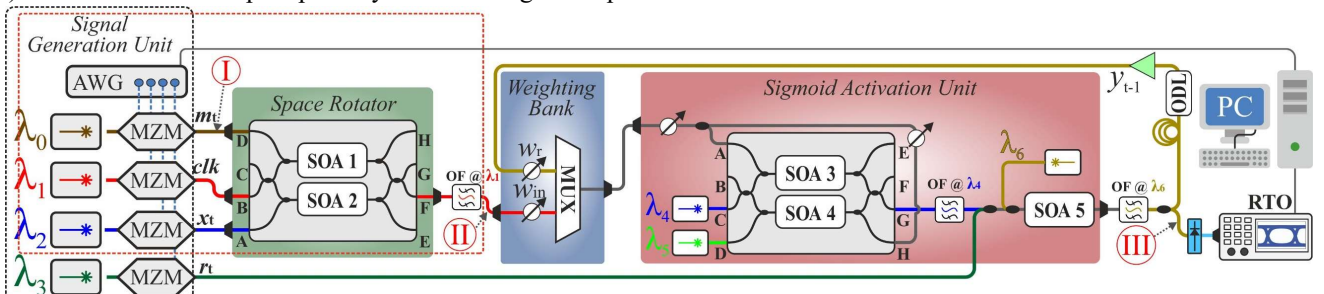

Fig. 3. Experimental setup used to validate the classification of 3-bit RZ & NRZ time-series through the proposed PRN as well as the functionality of the space rotator.

generating all the possible 3-bit sequences. Then, one of them was selected as the "targeted" bit pattern and the PRN was configured to detect that pattern. The bits were fed sequentially into the PRN, the loss function was applied only at the last training step (where the output of the neuron was read), while the PRN was trained using the back propagation through time method [32]. Note that the backpropagation through time algorithm was used for training the networks, which allows for efficiently training networks with more than one hidden layers, as well as for multiple recurrent steps [32]. The squared loss function was employed for training the neuron:

$$\mathcal{L} = \sum_{i=0}^{N} (y_{iT} - l_i)^2,$$

where y_{iT} denotes the output of the neuron at the T-th (final) step when presented with the i-th training bit sequence, l_i is the training target (1 for the desired pattern, 0 for the rest of them) and N is the number of samples in the batch. Note that a sample refers to a 3-bit pattern, while a batch refers to N such samples. Using large batches allows for accelerating the training process. For the conducted experiments, where the pattern of 3-bits was used, $N=9$ (all the samples were used for each iterations) and $T=3$ (the whole bit pattern was fed to the neuron before reading its output). The Adam algorithm with a learning rate of $\eta=0.0001$ (selected using line search to avoid slowing down the training process and ensure its smooth convergence) was used for the optimization [33], while the optimization ran for 10,000 iterations. It is worth noting that all the networks were trained using the PyTorch framework, in order to employ GPUs to accelerate the training process. However, instead of using regular neurons, we implemented the transfer functions of all the photonic components using the PyTorch framework, which allowed us to train the neurons by simulating their actual behavior when implemented on photonic hardware. Furthermore, the same training procedure can also be applied for larger bit sequences, since state-of-the-art GPUs and TPUs can support the training of models that require long sequences or multi-dimensional data.

B. Experimental setup of the space rotator & PRN

The experimentally implemented setup of the PRN is depicted in Fig. 3 and its performance was evaluated with every bit being represented by a 100-psec long optical pulse, both when an RZ as well as when an NRZ modulation scheme at 10Gb/s was employed. The Signal Generation Unit was responsible for generating the required signals that have been used for the experimental evaluation, with the sequence of 3-bit-long rotation mask m_t signals being imprinted on λ_0 and the sequence of 3-bit-long input vector x_t imprinted on λ_2. Moreover, an optical clock signal with 100-psec long optical pulses was generated at λ_1 that was responsible for carrying the result of the rotating function between the mask and the input time vector, while an additional signal carried by λ_3 had the role of the reset signal, denoting the end of every 3-bit input sequence. By employing binary signals for both the rotation mask and the input time vector, the output of the binary space rotator is equal to zero only when the corresponding bit of the rotation mask m_t and the time-series x_t are equal, yielding an

optical pulse at its output in any other case.

The space rotator was realized by means of a SOA-MZI that employed a CW optical beam at λ_1 as its input signal, an optical signal denoted as rotation mask m_t signal at λ_0 and an optical binary signal at λ_2 as its input time-series vector x_t. Both the m_t and x_t optical signals were launched as control signals into the respective SOAs of the two SOA-MZI branches through SOA-MZI ports A and D, respectively, with the presence of the mask m_t resulting in the successful operation of the SOA-MZI module as space rotating element whose output is governed by the equation (1). The space rotator output was imprinted on λ_1 that was subsequently launched into the weighting bank module, after being filtered in an Optical band-pass Filter (OF) with a 3dB-bandwidth of 0.8nm. This signal was then weighted by w_{in} in a variable optical attenuator prior being multiplexed with the optical recurrent signal in an optical wavelength multiplexer (MUX). The MUX output enters then the all-optical sigmoid activation function after being attenuated in a variable optical attenuator to realize the bias of the neuron and then getting split into 2 identical copies that are both injected into the SOA-MZI as control signals via the respective SOA-MZI ports A and H. The sigmoid activation unit is identical to the one demonstrated in [10] and has been realized by a deeply-saturated differentially-biased SOA-MZI followed by an SOA operating as Cross-Gain Modulation (XGM) gate in its small-signal gain regime. Two laser beams at λ_4 and λ_5 are forwarded into the SOA-MZI as the input and the auxiliary beam, respectively. The SOA-MZI output provides a clipped inverse copy of the control signal that gets imprinted on λ_4, which is subsequently combined with the reset signal carried by λ_3 and then jointly forwarded as a control signal into the SOA along with another CW laser beam at λ_6 that serves as the SOA input signal. The XGM operation carried out within the SOA yields an inverse copy of the signal carried by λ_4 to be imprinted on λ_6 and to emerge at the SOA output, resulting in this way to a SOA output that follows a sigmoid response with respect to the optical signal emerging at the MUX output.

The SOA output was then filtered by a 0.8nm OF centered at λ_6 and was split into 2 identical signals, with the first one being injected into a photodiode for being captured and analyzed in a Keysight DSOZ632A RTO with 33GHz bandwidth and 80GSa/s sampling rate. The second signal constituent was fed into a fiber-based feedback loop that comprises a fiber length of 61m and an optical delay line (ODL), so as to introduce a time delay of $D=T\times(N\times K-1)$ bits, with T denoting the bit period, K representing the number of bits contained within a signal period, and N being an integer. During the experimental evaluation of the RZ and the NRZ formats the K was 40, the T was equal to 300psec and 100psec, respectively, resulted in an N equal to 27 and 71, respectively. The delayed signal was then amplified by means of an Erbium-doped fiber amplifier (EDFA) and then was forwarded into the weighting bank, effectively forming the y_{t-1} recurrent neuron signal that was subsequently weighted by w_r via a variable optical attenuator prior being multiplexed with the respective weighted x_t signal in the optical multiplexer. Note that the photonic recurrent neuron comprising the weighting bank and the sigmoid

activation unit incorporated in the fiber loop was trained to perform successful recognition of the bit pattern $\tilde{x}_t$="111", having weight values equal to w_{in}=10.7dB and w_r=5.5dB, as suggested by the software-based training process. The configuration of attenuators has been realized with an error of ±0.01dB, determined by the respective resolution capabilities of the employed VOA modules.

C. Experimental results

The optically implemented functionality of the space rotator and the recurrent neuron when operated as individual units can be identified in Fig. 4, which depicts their optical outputs together with the respective inputs when injecting indicative RZ waveforms with 100-psec long pulses. Figure 4(a) shows an indicative waveform with analog pulse amplitude values that is inserted into the space rotator module as one of the two control signals, with the pulse amplitude value corresponding to the x_t input series vector. Figure 4(b) shows the respective optical output of the space rotator when using a constant DC power level as the second control signal of the space rotator, i.e. when a constant "1" is employed as the rotation mask. As can be clearly identified, the space rotator optical output yields the result of the $g(x_t) = (1 - x_t)$ mathematical expression, with the red arrows in Figs. 4(a) and 4(b) depicting two clear examples on how a pulse amplitude of x_t gets transformed onto a power level of $1-x_t$ at the output. The functionality of the recurrent neuron can be confirmed by means of Fig. 4(c) and 4(d), which illustrate an optical pulse stream used as incoming time-series vector into the photonic recurrent neuron and the corresponding optical pulse stream obtained at the neuron's output, respectively. As can be seen, the recurrent neuron can successfully and in real-time identify the presence of three successive '1's within a bit sequence that includes all possible
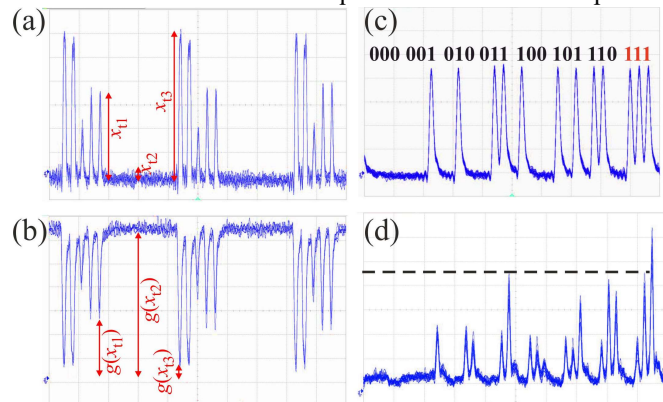

Fig. 4. Time traces from the experimental demonstration of the (a)-(b) space rotator as well as the PRN input (c) and output (d) employing a digital RZ waveform with 100psec pulses, respectively. (a)-(b) *y*-axis scale: (2mV/div) and *x*-axis scale:(1nsec/div), (c)-(d) *y*-axis scale: (2.9mV/div) and *x*-axis scale: (980psec/div).

3-bit long word combinations, emitting a clearly stronger optical pulse at the neuron output when the third successive "1" enters the neuron.

The experimental evaluation of the space rotator has been realized by feeding the space rotator with CWs at λ_0=1553.6nm, λ_1=1548.7nm, λ_2=1547.4nm and λ_3=1549.3nm. An analog RZ time-series signal x_t of four discrete power levels comprised of

100psec pulses, T=300psec symbol period and 4nsec signal period was injected into the port D of the space rotator having the role of time-series signal x_t. The rotation mask signal m_t="1" and the input signal has been realized by disabling the corresponding AWG channels to produce 2 separate CWs that have been forwarded into port A and B, respectively. The analog NRZ time-series signal x_t has been produced with T=100psec symbol period and 600psec signal period, having also four different power levels as in the RZ case. The optical peak power levels for the optical signals entering the space rotator were 2.7dBm, -0.2dBm and 3dBm, measured at the ports A, B and D, respectively. SOA1 and SOA2 were driven by a DC current of 280mA and 300mA, respectively.

Real-time time-series classification operation when cascading the recurrent neuron to the space rotator, as depicted in Fig. 3, and using 100-psec long optical pulses within a bit-period of T=300psec for the optical rotation mask m_t, input time-series x_t and input clock signals, are shown in Fig. 5(a)-(j). Figure 5(a) illustrates the bit time-series x_t that contains all the 8 different 3-bit pattern combinations between "000" and "111", while the clock and the rotation mask m_t signals have a periodic content of "1110" and "$X_1X_2X_30$", respectively. When a mask signal of m_t="000" is used, the output of the space rotator is identical to the input signal x_t of Fig. 5(a), as shown in Fig. 5(b), and the PRN has the role of identifying the incoming bit sequence of "111". The corresponding PRN output is illustrated in Fig. 5(c), clearly revealing that the highest amplitude output pulse emerges at the end of the "111" input time vector, indicating successful all-optical recognition of this specific bit pattern. The red line in Fig. 5(c) illustrates the respective software-obtained output of the software-trained PRN that provides almost a perfect matching with the corresponding experimentally obtained waveform. Figures 5(d)-(f) depict respective experimental results when the bit pattern "110" has to be identified by the PRN within the sequence of the input time vector. In this case, the input time-series of Fig. 5(d) is identical to the input sequence of Fig. 5(a), but the rotation mask used equals now m_t="001". In this way, the space rotator output, depicted in Fig. 5(e), has transformed the incoming "110" x_t bit pattern into a sequence of "111", which can be now again successfully recognized by the photonic recurrent neuron yielding the highest amplitude pulse across the entire PRN output sequence, as illustrated in Fig. 5(f). When the "100" input bit sequence has to be recognized, then a mask pattern of m_t="011" is used for transforming the incoming x_t time vector and the respective results are shown in Figs. 5(g)-(i). Figure 5(g) illustrates the sequence of the 8 different combinations of the input time vector x_t, while the space rotator output is illustrated in Fig. 5(h). As can be seen, the "100" input bit pattern has been successfully transformed into "111", which is then again successfully recognized by the recurrent neuron providing the highest amplitude optical pulse at the PRN output, as shown in Fig. 5(i).

Successful 3-bit string recognition has been also performed when using NRZ optical pulses at 10Gb/s, demonstrating the highest operational speed reported so far among recurrent neural network implementations. Figures 5(j)-(l) illustrate the
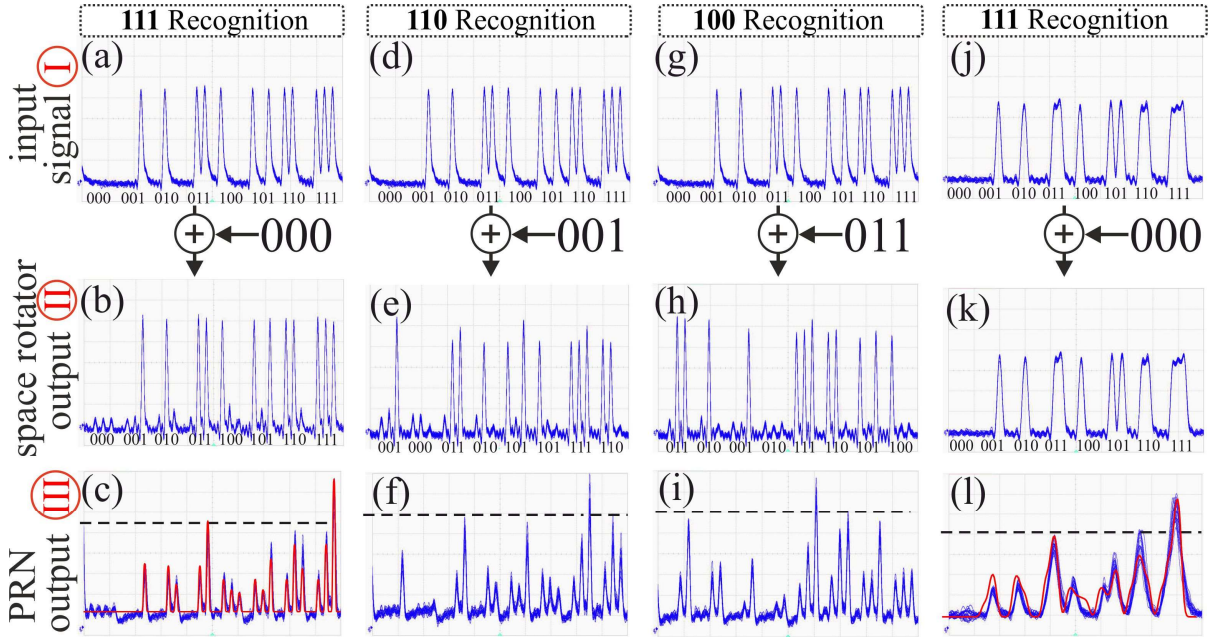
Fig. 5. Time traces from the experimental evaluation of 3-bit RZ time-series classification with space rotation mask: (a)-(c) m_t="000", (d)-(f) m_t="001" and (h)-(j) m_t="011". (k)-(m) present experimental results for a mask of m_t="000" using NRZ data. y-axis scale: (2.90mV/div), (a)-(j) x-axis scale:(980psec/div), (k)-(m) x-axis scale: (500psec/div).

obtained results for the NRZ bit sequence classification process when a mask of m_t="000" is employed, implying that the 3-bit sequence of "111" has to be identified by the PRN and the space rotator output will contain a bit pattern that is identical to the incoming time-series x_t, as shown in Fig. 5(k). Figure 5(l) illustrates the respective PRN output overlaid with the corresponding output of the software-based task execution, clearly revealing that the highest amplitude pulse above the dashed-line threshold is obtained at the end of the "111" sequence and that almost perfect matching with the software-obtained results is accomplished.

The wavelengths of the laser beams used for feeding the sigmoid activation unit were λ_4=1551.1nm, λ_5=1546.3nm and λ_6=1550.1nm. The optical peak power levels for the optical signals entering the space rotator during the PRN operation as 3-bit classifier were 2.7dBm, -1dBm and 1.2dBm, measured at the ports A, B and D, respectively. The corresponding peak power levels of the optical signals entering the SOA-MZI module of the sigmoid activation element were 7dBm, 2.2dBm, 3dBm and 6dBm at the SOA-MZI ports A, C, D and H, respectively. The reset signal, the output of SOA-MZI and the CW laser beam at λ_6 were launched into the SOA with a peak power of -7dBm, -6dBm and -9dBm, respectively. SOA1, SOA2 SOA3, SOA4 and SOA5 were driven by a DC current of 240mA, 260mA, 254mA, 278mA and 280mA, respectively.

The 3-bit classification accuracy for both the 300psec bit-period and the 10Gb/s NRZ bit sequences has been measured by capturing the respective PRN output time traces at the Real-Time Oscilloscope (RTO) and processing the captured information via a PC. The corresponding results are depicted in the bar charts of Fig. 5, with Fig. 6(a) showing the accuracy for all 8 possible bit combinations when an optical pulse stream with 300psec bit-period is employed and Fig. 6(b) illustrating the respective accuracy bars when a 10Gb/s NRZ optical bit

stream is used. Figure 6(a) reveals an average accuracy of 91.12% with a standard deviation of only 0.78%, with the lowest and highest values being 90% and 92.7% and corresponding to identified bit patterns of "110" and "111", respectively. Similar results were obtained for all possible 3-bit combinations also in the 10Gb/s NRZ case, as shown in Fig. 6(b), where the average accuracy was found to be 90.13% with
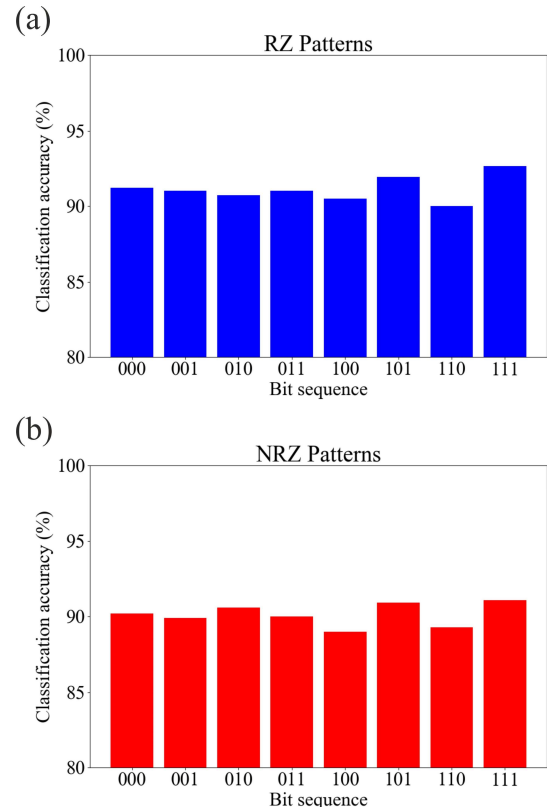

Fig. 6. Bar chart that depicts the measured accuracy for the 3-bit time-series classification task with (a) RZ and (b) NRZ patterns.

a standard deviation of only 0.68%. Lowest and highest accuracy values were measured to be 89% and 91.08%, corresponding to the 3-bit strings of "100" and "111", respectively. For both formats, the OSNR was equal to 11.2dB.

D. Accuracy measurement techniques

The PRN output signal was captured and analyzed in a Keysight DSOZ632A RTO with 33GHz bandwidth and 80GSa/s sampling rate. For the accuracy measurements in all possible 3-bit classification experiments, a total number of 80.000 samples was recorded by means of the RTO and was then subsequently inserted into an in-house developed Python software code. The software calculated the local maxima of each symbol resulting in its peak power value. This peak value has been compared with the threshold value that was estimated manually, taking into account the visual representation of the corresponding time trace. After the first waveform capture, another one has been performed in order to fine tune the threshold according to the number of errors that occurred above and below of it, targeting in equal distribution of errors. Thus, the final accuracy measurement has been performed on the third capture. It should be mentioned that the post processing of the results has been performed in order to set a threshold and measure in this way the accuracy of the classifier, without applying any offline training techniques similar to RC.

IV. DISCUSSION

Given that all components being required for the PRN layout can be available as integrated photonic modules, the proposed 3-bit time-series classifier can be fully compatible with available photonic integration platforms and can be eventually deployed as fully integrated photonic circuit, being capable of recognizing any incoming 3-bit pattern and ultimately yielding identification of a complete time series that comprises by 3-bit words. Figure 7(a) illustrates a potential integrated layout of the proposed architecture, where an 1-to-8 optical splitter is used at its front-end to broadcast 8 identical copies of the input 3-bit sequence to respective 8 PRN modules, with every PRN module being configured for operation with a different 3-bit rotation mask. Correlating then the outputs of all 8 PRN modules with the respective time slots where every PRN produces a successful classification outcome, the complete incoming time series can be identified and decoded. Figure 7(a) shows in more detail the case when the time series "001 111 101…" is injected into the 3-bit classification processor, depicting also the successful pulse generation at the PRN#2 output when the preceding "001" bit sequence enters the processor.

Figure 7(b) presents a close-up view of an integrable single PRN. The brown sections correspond to the passive components including tunable symmetric and asymmetric MZIs that are used for implementing the optical filtering and weighting functions required in the PRN. The yellow segments designate the areas where the active non-linear optical elements have to be incorporated for equipping the circuit with the mathematical operations required both in the space rotator and the sigmoid activation unit. In the case of employing the InP

monolithic integration platform [34], SOAs and SOA-MZIs can be directly employed also in the integrated version of the PRN and the 3-bit classifier, presumably yielding a PRN that consumes more than 3 Watts when taking into account that each SOA requires close to 600mW. A more compact and lower energy consumption chip-scale layout can be, however, accomplished if migrating to the silicon photonic integration platform and adopting InP-on-Si photonic crystal devices for replacing the SOA-based nonlinear functionalities, taking advantage of the footprint and energy efficiency of silicon photonics over InP monolithically integrated passive and active phase tuning structures as well as of the InP-on-Si photonic crystals (PhC) over SOAs. InP-on-Si photonic crystal switch and injection-locked laser technology [35] can serve as the space rotator and sigmoid activation elements, respectively, having been demonstrated to perform at 10Gb/s with energy efficiencies as low as 15.1fJ/bit and 13.5fJ/bit, respectively. This suggests that a complete PRN where the space rotator is accomplished via a single InP-on-Si PhC switch and a sigmoid activation module via an InP-on-Si switch followed by an injection locked InP-on-Si laser unit could possibly operate at 10Gb/s with a total power consumption of only ~450uW, yielding an overall energy efficiency of only ~225 fJ/bit.

The deployment of more sophisticated classification tasks will require the implementation of a photonic RNN with gated mechanism as has already been demonstrated in [36]. This layout resembles the Minimal Gated Unit (MGU) functionality of [37], allowing the deployment of financial prediction tasks as well as text recognition and natural language processing with an accuracy similar to LSTMs and GRUs [36]. Regarding the recurrent iterations of each neuron, we believe that the regenerative properties of the differentially-biased scheme will allow for 6 iterations at least. The data rate of the SOA-based PRN can be further increased to 40Gbit/s by operating the SOA-MZIs in a differentially biased configuration [38] and all the SOAs being operated in the deeply saturated regime [39], [40]. Finally, the number of inputs for each SOA-based neuron can scale up to 32 [41].
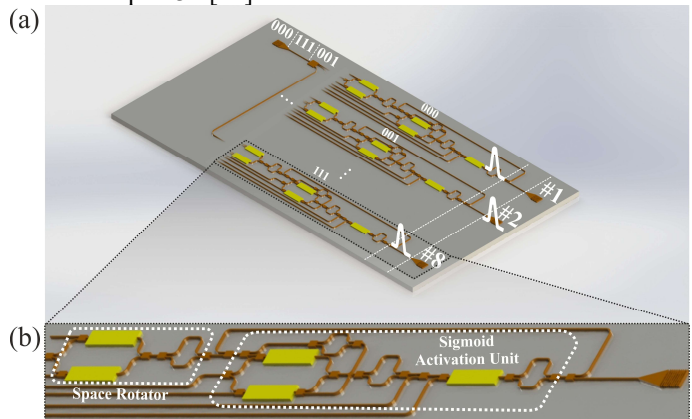
Fig. 7. Schematic layout of a possible integrated version of (a) an all-optical 3-bit classification processor and (b) an individual integrated PRN.

V. CONCLUSION

This work presents time-series classification at the highest operational speed reported so far without employing any off-

line post-processing underlying on novel a 10Gb/s photonic recurrent neuron. We demonstrate experimentally successful 3-bit optical time-series recognition at 3.3 and 10Gb/s for RZ and NRZ formats with an average accuracy of 91.12% and 90.13%, respectively. The PRN relies on an all-optical space rotating module followed by an all-optical sigmoid recurrent neuron, enabling successful classification process with optical power signals and weight values that extend only along the range of positive values, overcoming in this way the extra complexity and circuitry associated with negative number representation in all-optical neural network deployments [4]. This solution offers the first ever reported photonic recurrent neural network application over time-series vectors offering real-time classification, taking the first important step towards the deployment of more sophisticated recurrent layouts such as LSTMs and GRUs at much higher operational speeds compared to their electronic counterparts.

REFERENCES

[1] H.-T. Peng, M. A. Nahmias, T. F. de Lima, A. N. Tait, and B. J. Shastri, "Neuromorphic Photonic Integrated Circuits," *IEEE J. Sel. Top. Quantum Electron.*, vol. 24, no. 6, pp. 1–15, 2018, doi: 10.1109/JSTQE.2018.2840448.

[2] M. A. Nahmias, T. F. De Lima, A. N. Tait, H. T. Peng, B. J. Shastri, and P. R. Prucnal, "Photonic Multiply-Accumulate Operations for Neural Networks," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 1, p. 1, 2020, doi: 10.1109/JSTQE.2019.2941485.

[3] A. R. Totovic, G. Dabos, N. Passalis, A. Tefas, and N. Pleros, "Femtojoule per MAC Neuromorphic Photonics: An Energy and Technology Roadmap," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 5, pp. 1–15, Sep. 2020, doi: 10.1109/JSTQE.2020.2975579.

[4] A. N. Tait *et al.*, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017, doi: 10.1038/s41598-017-07754-z.

[5] A. N. Tait *et al.*, "Feedback control for microring weight banks," *Opt. Express*, vol. 26, no. 20, p. 26422, Oct. 2018, doi: 10.1364/OE.26.026422.

[6] A. N. Tait, A. X. Wu, T. Ferreira de Lima, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Two-pole microring weight banks," *Opt. Lett.*, vol. 43, no. 10, p. 2276, May 2018, doi: 10.1364/OL.43.002276.

[7] Z. Cheng, C. Ríos, W. H. P. Pernice, C. David Wright, and H. Bhaskaran, "On-chip photonic synapse," vol. 2, no. September, pp. 1–7, 2017, [Online]. Available: http://advances.sciencemag.org/.

[8] S. Abel *et al.*, "Silicon photonics integration technologies for future computing systems," in *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, Jul. 2019, pp. 1–3, doi: 10.23919/PS.2019.8818051.

[9] B. Shi, N. Calabretta, and R. Stabile, "Deep Neural Network Through an InP SOA-Based Photonic Integrated Cross-Connect," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 1, pp. 1–11, Jan. 2020, doi: 10.1109/JSTQE.2019.2945548.

[10] 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, p. 9620, 2019, doi: 10.1364/OE.27.009620.

[11] M. Miscuglio *et al.*, "All-optical nonlinear activation function for photonic neural networks [Invited]," *Opt. Mater. Express*, vol. 8, no. 12, p. 3851, Dec. 2018, doi: 10.1364/OME.8.003851.

[12] A. N. Tait *et al.*, "Silicon Photonic Modulator Neuron," *Phys. Rev. Appl.*, vol. 11, no. 6, pp. 1–16, 2019, doi: 10.1103/PhysRevApplied.11.064043.

[13] C. Huang *et al.*, "Programmable Silicon Photonic Optical Thresholder," *IEEE Photonics Technol. Lett.*, vol. 31, no. 22, pp. 1834–1837, 2019, doi: 10.1109/LPT.2019.2948903.

[14] J. George *et al.*, "Electrooptic Nonlinear Activation Functions for Vector Matrix Multiplications in Optical Neural Networks," in *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*, 2018, p. SpW4G.3, doi: 10.1364/SPPCOM.2018.SpW4G.3.

[15] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics*, vol. 11, no. 7, pp. 441–446, Jun. 2017, doi: 10.1038/nphoton.2017.93.

[16] G. Mourgias-Alexandris *et al.*, "Neuromorphic Photonics With Coherent Linear Neurons Using Dual-IQ Modulation Cells," *J. Light. Technol.*, vol. 38, no. 4, pp. 811–819, Feb. 2020, doi: 10.1109/JLT.2019.2949133.

[17] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019, doi: 10.1038/s41586-019-1157-8.

[18] K. Vandoorne *et al.*, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nat. Commun.*, vol. 5, pp. 1–6, 2014, doi: 10.1038/ncomms4541.

[19] M. Freiberger, S. Sackesyn, C. Ma, A. Katumba, P. Bienstman, and J. Dambre, "Improving Time Series Recognition and Prediction With Networks and Ensembles of Passive Photonic Reservoirs," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 1, pp. 1–11, Jan. 2020, doi: 10.1109/JSTQE.2019.2929699.

[20] K. Vandoorne, J. Dambre, D. Verstraeten, B. Schrauwen, and P. Bienstman, "Parallel reservoir computing using optical amplifiers," *IEEE Trans. Neural Networks*, vol. 22, no. 9, pp. 1469–1481, 2011, doi: 10.1109/TNN.2011.2161771.

[21] F. Laporte, A. Katumba, J. Dambre, and P. Bienstman, "Numerical demonstration of neuromorphic computing with photonic crystal cavities," *Opt. Express*, vol. 26, no. 7, p. 7955, 2018, doi: 10.1364/OE.26.007955.

[22] H. Bagherian, S. Skirlo, Y. Shen, H. Meng, V. Ceperic, and M. Soljacic, "On-Chip Optical Convolutional Neural Networks," pp. 1–18, 2018, [Online]. Available: http://arxiv.org/abs/1808.03303.

[23] P. R. Prucnal, B. J. Shastri, T. Ferreira de Lima, M. A. Nahmias, and A. N. Tait, "Recent progress in semiconductor excitable lasers for photonic spike processing," *Adv. Opt. Photonics*, vol. 8, no. 2, p. 228, 2016, doi: 10.1364/AOP.8.000228.

[24] N. Passalis, G. Mourgias-alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, "Training Deep Photonic Convolutional Neural Networks with Sinusoidal Activations," no. Dl, pp. 1–10.

[25] N. Passalis, G. Mourgias-Alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, "Variance Preserving Initialization for Training Deep Neuromorphic Photonic Networks with Sinusoidal Activations," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 1483–1487, doi: 10.1109/ICASSP.2019.8682218.

[26] M. Y.-S. Fang, S. Manipatruni, C. Wierzynski, A. Khosrowshahi, and M. R. DeWeese, "Design of optical neural networks with component imprecisions," *Opt. Express*, vol. 27, no. 10, p. 14009, 2019, doi: 10.1364/oe.27.014009.

[27] N. C. Harris *et al.*, "Linear programmable nanophotonic processors," *Optica*, vol. 5, no. 12, p. 1623, Dec. 2018, doi: 10.1364/OPTICA.5.001623.

[28] K. Vandoorne *et al.*, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nat. Commun.*, vol. 5, no. 1, p. 3541, May 2014, doi: 10.1038/ncomms4541.

[29] F. Akopyan *et al.*, "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 34, no. 10, pp. 1537–1557, 2015, doi: 10.1109/TCAD.2015.2474396.

[30] S. B. Furber *et al.*, "Overview of the SpiNNaker system architecture," *IEEE Trans. Comput.*, vol. 62, no. 12, pp. 2454–2467, 2013, doi: 10.1109/TC.2012.142.

[31] N. P. Jouppi *et al.*, "In-Datacenter Performance Analysis of a Tensor Processing Unit," *ACM SIGARCH Comput. Archit. News*, vol. 45, no. 2, pp. 1–12, 2017, doi: 10.1145/3140659.3080246.

[32] P. J. Werbos, "Backpropagation Through Time: What It Does and How to Do It," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990, doi: 10.1109/5.58337.

[33] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, doi: arXiv:1412.6980.

[34] F. M. Soares *et al.*, "InP-Based Foundry PICs for Optical Interconnects," *Appl. Sci.*, vol. 9, no. 8, p. 1588, 2019, doi: 10.3390/app9081588.

[35] T. Alexoudi *et al.*, "III-V-on-Si Photonic Crystal Nanocavity Laser

Technology for Optical Static Random Access Memories," *IEEE J. Sel. Top. Quantum Electron.*, vol. 22, no. 6, pp. 295–304, 2016, doi: 10.1109/JSTQE.2016.2593636.

[36]   G. Mourgias-Alexandris, G. Dabos, N. Passalis, A. Totovic, A. Tefas, and N. Pleros, "All-Optical WDM Recurrent Neural Networks With Gating," *IEEE J. Sel. Top. Quantum Electron.*, vol. 26, no. 5, pp. 1–7, Sep. 2020, doi: 10.1109/JSTQE.2020.2995830.

[37]   G. Zhou, J. Wu, C.-L. Zhang, and Z.-H. Zhou, "Minimal Gated Unit for Recurrent Neural Networks," Mar. 2016, [Online]. Available: http://arxiv.org/abs/1603.09420.

[38]   M. Spyropoulou *et al.*, "40 Gb/s NRZ wavelength conversion using a differentially-biased SOA-MZI: Theory and experiment," *J. Light. Technol.*, vol. 29, no. 10, pp. 1489–1499, 2011, doi: 10.1109/JLT.2011.2134832.

[39]   G. Mourgias-alexandris, A. Tsakyridis, T. Alexoudi, K. Vyrsokinos, and N. Pleros, "Optical Thresholding Device with a Sigmoidal Transfer Function," *Photonics Switch. Comput.*, pp. 6–8, 2018.

[40]   A. Tsakyridis *et al.*, "Theoretical and Experimental Analysis of Burst-Mode Wavelength Conversion via a Differentially-Biased SOA-MZI," *J. Light. Technol.*, vol. 38, no. 17, pp. 4607–4617, 2020, doi: 10.1109/JLT.2020.2995471.

[41]   M. Moralis-Pegios, N. Terzenidis, G. Mourgias-Alexandris, K. Vyrsokinos, and N. Pleros, "A 1024-Port Optical Uni-and Multicast Packet Switch Fabric," *J. Light. Technol.*, vol. 37, no. 4, pp. 1415–1423, 2019, doi: 10.1109/JLT.2019.2894510.