All-optical recurrent neural network with sigmoid activation function

George Mourgias-Alexandris, George Dabos, Nikolaos Passalis, Anastasios Tefas, Angelina Totovic and Nikos Pleros
[*] *Department of Informatics, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece*
mourgias@csd.auth.gr

Abstract: We demonstrate experimentally, the first all-optical recurrent-neuron with a sigmoid activation function and four WDM-inputs with 100psec pulses. The proposed neuron geared up a neural-network for financial prediction-tasks exhibiting an accuracy of 42.57% on FI-2010.

1. Introduction

Neuromorphic photonics are highly anticipated to revolutionize future computational systems by transferring their low-power consumption, time-of-flight latency and THz bandwidth credentials into neuromorphic hardware implementations [1]. In this context, research efforts have mainly focused on the deployment and demonstration of weighting banks [2], activation functions [3] linear neuron layouts [4] and novel training frameworks tailored to neuromorphic photonic configurations [5], with the main emphasis lying, however, on feed-forward neuron architectures [4], [6]. Although feed-forward architectures can indeed yield powerful configurations for a large set of applications, time-series analysis, classification and prediction tasks are well-known to comprise the stronghold of Recurrent-Neural-Networks (RNNs), which form also the backbone of more sophisticated Long-Short-Term-Memory and Gated-Recurrent-Unit (GRU) architectures with unmatched processing capabilities. The implementation of RNNs directly in the optical domain would probably allow for performing time-series analysis at much higher line-rates compared to respective electronic RNNs, allowing for significant latency advantages in processing time sequences. However, only a small number of deployments towards all-optical RNNs have been reported so far, with the main demonstrations comprising simpler autaptic circuits either for spiking [7] or non-spiking neural networks [8]. Turning autaptic circuitry into a full-scale Recurrent Neuron (RN) requires, however, the extra complexity of a multi-input linear neuron stage, which has not been reported so far. At the same time, migrating to all-optical RNs would benefit from the use of a sigmoid activation function that is much more commonly encountered in conventional AI training models for RNNs, offering significant benefits in terms of compatibility with the rich available Deep Learning training framework for time series processing tasks [9].

In this paper, we experimentally demonstrate for the first time, to the best of our knowledge, an all-optical RN with a sigmoid activation function using a WDM-based four-input linear neuron followed by a SOA-based all-optical sigmoid activation module [10] incorporated in a recurrent layout. The performance of the photonic RN has been experimentally evaluated with 100psec optical pulses with the experimental results validating the recurrent behavior of the proposed RN for different levels of the recurrent weight values and with effective thresholding for timely de-synchronized input signals with a power level contrast of up to 8dB. The potential of the photonic sigmoid RN module to penetrate time series processing functionalities is then outlined by successfully training an RNN architecture that employs the proposed all-optical RN for prediction tasks with the FI-2010 financial data set.

2. All-optical Sigmoid Recurrent Module: Experimental Setup and Results

Figure 1(a) illustrates the proposed RN that employs a recently demonstrated SOA-based optical sigmoid activation unit [10] being nested into a feedback delay loop. The principle of operation of the proposed RN is governed by the equation: $y_t = f(\sum(w_i x_t^i) + w_r y_{(t-1)})$ for $i=0\ldots M$, where M is the number of inputs, $f(x)$ is the transfer function of the sigmoid activation unit (with Fig.1(b) depicting the experimentally obtained response together with the theoretical sigmoid fitting [10]), x_t^i denotes the i-th input signal at time instant t, w_i is the weight enforced to each x_t^i, $y_{(t-1)}$ is the 1-bit delayed version of the RN output $y_{(t)}$ at time instant t and w_r the weight enforced within the recurrency loop at y_t. The experimental setup shown in Fig.1(a) uses only a single x_t input signal for i=1 for validating initially the successful recurrent operation prior scaling this into a full-scale RN with multiple input signals. An input periodic electrical signal x_t^1 with a period of K=4 symbols, a symbol-period of T=400psec and a pulse duration of 100psec was generated by an Arbitrary Waveform Generator (AWG) and was converted into a respective optical x_t^1 via a $LiNbO_3$ optical modulator that was powered by a CW optical beam at λ_0. This signal was then weighted by a factor of w_1 using a Variable-Optical-Attenuator (VOA) before entering the activation unit as the SOA-MZI control signal. More details about the detailed layout and the principle of operation of the SOA-based sigmoid activation module

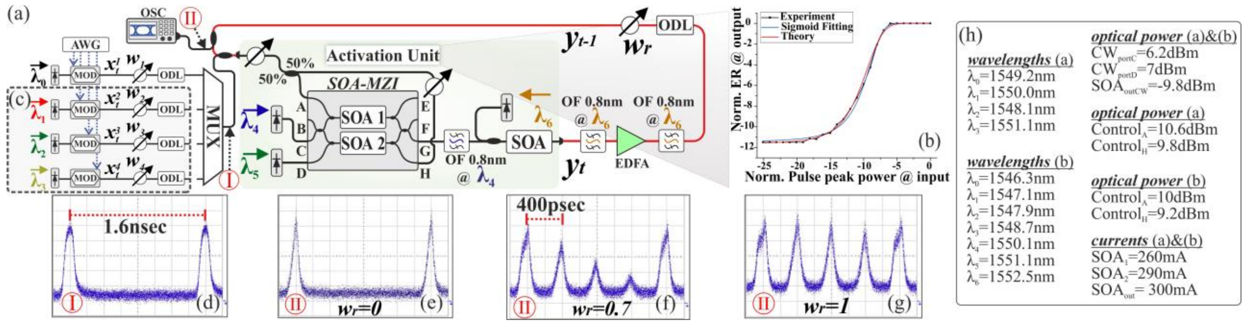


Fig. 1. (a) Experimental setup used to valid the recurrent capabilities of the proposed RN, (b) Sigmoid transfer function of the activation unit. (c) 3-input weighted signals. Experimental time traces (200psec/div) obtained for the single-λ layout for the at the (d) input signal, and (e)-(g) output of RN for attenuation factors of: (e) w_r=0, (f) w_r=0.7 and (g) w_r=1. *y*-axis scale: (3.5mV/div). (h) Panel summarizing operational conditions for both single-λ and WDM RN experiments.

can be found in [10]. The activation module output signal carried by λ_6 enters then a 61m-long fiber loop that weights the output signal by a w_r factor implemented by a VOA and enforces a time delay equal to $T\times(N\times K\text{-}1)$, with N being an integer. In this way, the recirculating signal that gets combined and added with the next input symbol $w_1x_t^1$ via a fiber coupler prior entering again the activation stage carries the $w_r y_{(t\text{-}1)}$ information. An Erbium-Doped Fiber Amplifier (EDFA) and Optical Bandpass Filters (OF) were used to amplify and filter the recurrent signal. To validate the recurrent operation of the proposed neuron, several different weighting settings (w_r) were investigated while keeping the input weight w_1 constantly set to 1 (zero attenuation) for an input periodic sequence of "1000". The time trace of the input signal is illustrated in Fig. 1(d). Figure 1(e), illustrates the neuron output for the highest possible attenuation within the fiber loop, so that $w_r=0$. Decreasing the attenuation level by 4dB that yields a $w_r=0.7$, three pulses with gradually decaying peak powers are formed within every signal period and every 400psec, as shown in Fig. 1(f), confirming the successful operation of the recurrency module. Finally, when no attenuation is applied to the recirculating signal and w_r=1, these three pulses get almost power equalized with the initial input pulse, as can be clearly seen in Fig. 1(f). The operational conditions of the experiment are summarized in the panel shown in Fig. 1(h).

3. All-optical WDM Sigmoid Recurrent Neuron

The all-optical sigmoid recurrent module of Fig. 1(a) was transformed into a 4-input WDM RN by connecting the 3-input signal generation setup, shown in Fig. 1(c), to the respective 3 free ports of the 4:1 multiplexer unit (MUX). These three additional input signals carried at λ_1, λ_2 and λ_3 were individually modulated with the same periodic electrical signal x_t^i, that was used for modulating the first input signal employed also in the single-λ setup for the experimental validation of the recurrent module. ODLs were used prior the multiplexing operation in order to realize the timely-synchronized and de-synchronized addition of the four input signals. Figure 2(a) illustrates the spectrum of the resulting WDM sum obtained at the MUX output. In the case of timely-synchronized operation, the input signal is depicted in Fig. 2(b), while the RN output is shown in Figs. 2(c)-(e) for different weighting factors w_r. In all cases experimental results validate the recurrent behavior of the neuron. Figure 2(f) illustrates the input signal when the $w_3x_t^3$ signal gets de-synchronized and delayed by $2T$ (800psec) with respect to the remaining three input signals. Figs. 2(g)-(i) show the obtained results for different w_r weighting factors and different operational regimes of the sigmoid activation response, depicting that the RN can yield all possible output states between an output state that is

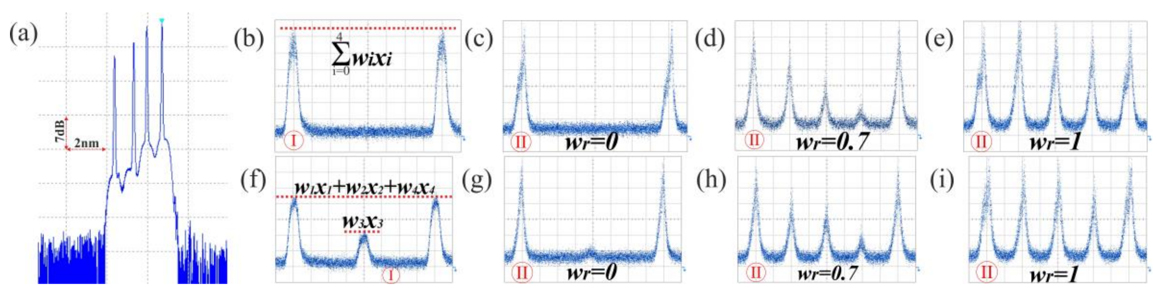


Fig. 2. (a) Spectrum of the WDM weighted input signals with a power contrast up to 8 dB. Time traces (200psec/div) for the the-(b) timely-synchronized WDM input as well as the RN output with an attenuation factor of: (c) w_r=0, (d) w_r=0.7 and (e) w_r=1. Respective results after inducing a delay of two time-slots to the $w_t^3x_t^3$ signal are illustrated in (i)-(f). *y*-axis scale: (3.5mV/div)

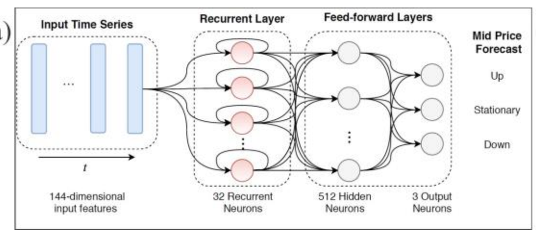
identical to the input signal [Fig. 2(g)], an output state with gradually decaying recirculating pulses formed within the signal period [Fig. 2(h)] and, finally, a power-equalized pulse stream with a period equal to *T*. The operational conditions of the experiment are summarized in the panel shown in Fig. 1(h).

4. Training and inference of an RNN

The proposed recurrent neuron was trained with standard deep learning tools using the high-frequency limit order book dataset ("FI-2010") [11], in the context of financial prediction tasks. The complete architecture of the trained network, considering 32 RNs in total, is depicted in Fig. 4(a). The optimization ran for 20 epochs using the RMSprop algorithm with a learning rate of $\eta=10^{-4}$ [12]. The photonic models were appropriately initialized by taking into account the behavior of the photonic activation function [5].

The results for the FI-2010 dataset are presented in Fig 4(b). The performance of the proposed photonic recurrent neuron was also compared to an MLP baseline, as well as to an RNN model that uses the ideal sigmoid activation function instead of the photonic one. The precision score, which is defined as the number of correctly classified instances (true positives) over the total number of instances that the network classified in one specific class (true and false positives) is reported. The proposed Photonic RNN achieves an average precision score of 42.57%, significantly outperforming the MLP model (39.26%) while being also competitive to an RNN using regular sigmoid activations (42.60%).

(a)



(b)

Model	Avg. Precision
MLP	39.26 ± 0.74 %
RNN	42.60 ± 1.18 %
Photonic RNN	42.57 ± 1.66 %

Fig. 4. (a) Recurrent neural network architecture and (b) Evaluation results for different neural network architectures.

5. Conclusion

We demonstrated experimentally the first all-optical recurrent neuron with a sigmoid activation function exploiting four WDM weighted input signals with 100psec optical pulses. The proposed neuron was trained using the FI-2010 data set exhibiting an accuracy of 42.57%. The 4-input all-optical RN provides a computational efficiency of 40 GMAC/sec and can presumably scale up to 32-input RN layouts by utilizing the proven capabilities of SOA-based nonlinear modules to perform over a broadband 32-channel regime [13], which would allow for 0.32 TMAC/sec computational efficiencies and 6.25pJ/bit energy consumption. Future on-chip implementation could also replace the fiber-based with integrated waveguide-based 1-bit recurrency, yielding energy efficiencies as low as 1.2fJ/bit by replacing the SOA devices with InP-on-Si Photonic Crystals that can provenly offer the sigmoid response [14].

6. Acknowledgements

EU-H2020 projects MASSTART (825109), PLACMOS (780997) & PLASMONIAC (871391).

7. References

[1] H. T. Peng et al., "Neuromorphic Photonic Integrated Circuits," IEEE J. Sel. Top. Quantum Electron., vol. 24, no. 6, pp. 1–15, 2018.

[2] A. N. Tait et al., "Demonstration of WDM weighted addition for principal component analysis," Opt. Express, vol. 23, no. 10, p. 12758, 2015.

[3] 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.

[4] G. Mourgias-alexandris et al., "Neuromorphic Photonics with Coherent Linear Neurons using dual-IQ modualtion cells," JLT, 2019.

[5] N. Passalis et al., "Training Deep Photonic Convolutional Neural Networks with Sinusoidal Activations," no. Dl, pp. 1–10.

[6] Y. Shen et al., "Deep learning with coherent nanophotonic circuits," Nat. Photonics, vol. 11, no. 7, pp. 441–446, Jun. 2017.

[7] H. Peng et al., "Autaptic Circuits of Integrated Laser Neurons," in *Conference on Lasers and Electro-Optics*, paper SM3N.3.

[8] A. N. Tait et al., "Silicon Photonic Modulator Neuron," Phys. Rev. Appl., vol. 10, no. 1, p. 1, 2019.

[9] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network." 2018.

[10] G. Mourgias-Alexandris et al., "An all-optical neuron with sigmoid activation function," Opt. Express, vol. 27, no. 7, p. 9620, 2019.

[11] A. Ntakaris et al., "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," J. Forecast., vol. 37, no. 8, pp. 852–866, 2018.

[12] T. Tieleman and G. Hinton, "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude," COURSERA: Neural networks for machine learning, vol. 4, no. 2, pp. 26–31, 2012.

[13] M. Moralis-Pegios et al., "A 1024-Port Optical Uni- and Multicast Packet Switch Fabric," JLT, vol. 37, no. 4, pp. 1415–1423, Feb. 2019.

[14] 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, Nov. 2016.