Extreme Learning Machines based on Optical Frequency Combs

A. Lupo,[1] M. Zajnulina,[1] and S. Massar[1]

[1] Laboratorie d'Information Quantique, CP 224, Université libre de Bruxelles, Av. Roosevelt 50, 1050, Bruxelles, Belgium

Artificial Neural Networks (ANN) are powerful neuromorphic, i.e. brain-inspired, algorithms already successfully deployed in a plethora of fields including computer vision and natural language processing. Extreme Learning Machines (ELM) are a type of ANNs suitable for classification tasks. Contrary to traditional ANNs, ELMs have mostly fixed, untrained, connections: only the output weights are trained, making the network easier to implement on non-electronic substrates. We developed two schemes for photonic ELMs based on frequency combs, where neurons are encoded in complex amplitudes of comb lines. A Programmable Spectral Filter (PSF) encodes input information by setting the initial line amplitudes and/or phases. The subsequent mixing of encoded information occurs via interference between comb lines, which can be achieved either by applying a phase modulation or by propagating the comb through a nonlinear Kerr medium such as optical fiber. The output weights are applied by a second PSF, attenuating each comb line proportionally to the required weight. The weighted comb is then measured by a photodiode, introducing a quadratic nonlinearity. A preliminary comparison of the two different schemes of information mixing shows the superiority of the Kerr-medium approach.

Introduction

Human activities generate an increasing demand for computational power that cannot be satisfied by traditional digital computers anymore. Electronic chip miniaturization, following the Moore law, will eventually stagnate, mainly due to quantum and thermal limits [1]. Also, traditional digital computing architectures, despite being relatively easy to deploy, constitute a bottleneck in terms of energy and computational time consumption when it comes to the processing of big amounts of data.

The human brain is known to outperform traditional computers on tasks such as speech or visual data recognition. This suggests that for certain problems, brain-inspired algorithms and/or architectures can be advantageously used for data processing. In fact, non-digital brain-inspired algorithms, in particular Artificial Neural Networks (ANNs), have been already successfully deployed in the above-mentioned tasks, and more [2]. ANNs process information and approximate the desired output by transforming the input data as it propagates in a network composed of trainable weighted connections and nonlinear activations.

Extreme Learning Machines (ELMs, Fig. 1a) are a subclass of feedforward ANN algorithms in which most of the connections remain untrained [3]. Only output connections are trained to approximate the desired behavior. This approach has two main advantages. First, it simplifies and speeds up the training procedure (which often reduces to a linear regression problem). Second, it allows for the network to be easily implemented in unconventional computational substrates. ELM networks are organized into three layers: the input, hidden, and output layer.

Many physical systems that exhibit nonlinear behavior can be exploited as ELM (Fig. 1b). To this end, initial information is injected into the system via the input layer using a

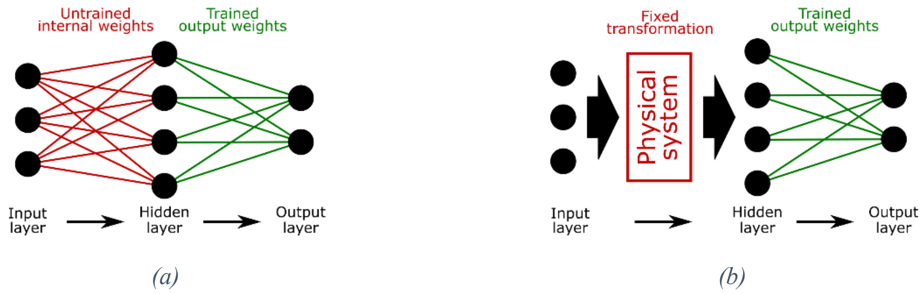

Figure 1. (a) The scheme of an Extreme Learning Machine. The number of neurons per layer is arbitrary. Connections between input and hidden layers (red) are fixed, while the connections between hidden and output layers (green) are trained. (b) The scheme of an Extreme Learning Machine implemented through an arbitrary physical system. The input layer is injected in the system and gets transformed in a fixed way (red), according to its internal rules. The state assumed by the system in response to the input is considered to represent the hidden layer. Multiplication by trained output weights (green) is performed in the readout layer.

suitable perturbation mechanism. The response to the perturbation is considered to be the hidden layer. The way in which perturbations are applied and the way in which responses are measured depend on the parameters of the system. In photonic systems, the perturbation could be a modulation of light, while the response could be given by a measurement of amplitudes or intensities. Hence, the physical system performs the transformation from input layer to hidden layer. A properly designed readout mechanism can measure the system response (hidden layer) and perform multiplication by trained output weights, thus generating the output layer of the ELM. Multiple photonics-based implementations of ELMs have been reported, which exploit propagation of light in free space [4], in multimode fibers [5], in time multiplexing systems [6], and more.
We presented a scheme of a photonic ELM based on wavelength multiplexing. In this scheme, each neuron is encoded in the amplitude of a line of an Optical Frequency Comb (OFC). The input layer is generated by properly shaping the OFC through a spectral filter, and the (untrained) information mixing occurs via OFC line interference stimulated by periodic phase modulation [7]. In the present work, we introduce a novel mixing method that exploits the Kerr nonlinearity of a standard telecom fiber. In the next section, we describe the experimental setup and the experimental procedure; then, we present and discuss the results; finally, we summarize our work in the conclusion.

Methods

The experimental setup is represented in Fig. 2. An OFC is generated by periodically phase modulating a monochromatic CW laser radiation. The laser wavelength, that determines the center of the OFC, is $\lambda_0 = 1554.6nm$. The periodic phase modulation is generated by a $LiNbO_3$ Phase Modulator (PM1) driven by a RF signal (power $P_1 = 32.7dBm$ and frequency $\Omega/2\pi = 16.969GHz$). The span of the OFC depends on the RF signal intensity P_1. We chose P_1 to generate 31 usable comb lines with spectral spacing of Ω. Input information is encoded in the central lines of the OFC by a Programmable Spectral Filter (PSF1, II-VI WaveShaper). The inputs for each neuron are always normalized in the range of values $[0,1]$, then linearly remapped in the range of attenuations $[-30dB, 0dB]$. The modulated OFC constitutes the input layer of our ELM. Our system is designed to test three different schemes for information mixing: phase modulation, Kerr nonlinearity in fiber, and both. The mixing based on phase modulation

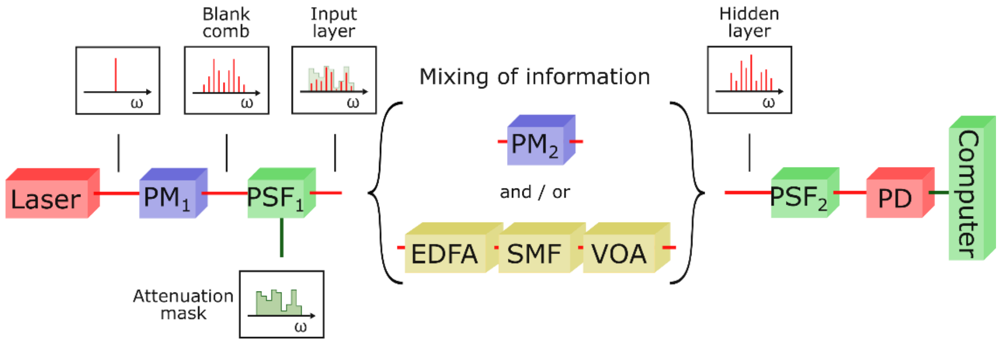

Figure 2: Experimental realization of a photonic Extreme Learning Machine. PM: Phase Modulator, PSF: Programmable Spectral Filter; SMF: Single Mode Fiber spool; VOA: Variable Optical Attenuator; PD: Photodiode

is obtained by propagating the OFC through another $LiNbO_3$ Phase Modulator (PM2) that is driven at the frequency Ω and RF power $P_2 = 21.4dBm$. The phase modulation generates interference between comb lines, and, thus, provides the required mixing of information [7]. The mixing based on the Kerr effect is obtained by amplifying the signal up to $27dBm$ through an Erbium-Doped Fiber Amplifier (EDFA) and then propagating it along a standard telecom Single Mode Fiber (SMF) with a length of $540m$. In the fiber, the information is mixed due to the Kerr-effect-based self- and cross-phase modulation. A Variable Optical Attenuator (VOA), placed immediately after SMF, reduces the intensity to a power level safe for the rest of the system. When both mixing methods are tested at the same time, the phase modulation is applied after the propagation in fiber. The OFC obtained after information mixing encodes the hidden layer of the ELM. The last part of the optical circuit is the readout layer implemented via a second Programmable Spectral Filter (PSF2, II-VI WaveShaper) and a Photodiode (PD). PSF2 is employed as a programmable bandpass filter. The OFC lines are measured individually, and their intensities are recorded on a computer by tuning the filter. The set of OFC line intensities constitutes the hidden layer. Multiplication by optimal output weights is performed on the computer. The multiplication by output weights can also be done all-optically as described in [7].

The proposed system is tested using three benchmark classification tasks (Iris classification [8], Wine classification [9], and Heart disease classification [10]). Their performances are evaluated in terms of accuracy, i.e. percentage of correct classifications. For each task, we split the set of available samples, using 70% of the samples to train the output weights and the remaining 30% to evaluate the accuracy. The output weights are trained using a ridge regression model (cf. [7]).

Results and discussion

We compare three different mixing schemes (phase modulation, Kerr nonlinearity, and both) using three different classification tasks (Wine, Iris and Heart disease classifications). Also, we compare our experimental results with the performance obtained by a software ELM having a similar topology (25 hidden neurons, quadratic nonlinearity). To acquire better statistics, the accuracy values are measured as the average score calculated over 10^3 random repartitions of the samples into training and testing sets. Results are reported in Table 1.

The presence of Kerr nonlinearity clearly boosts the performance with respect to the presence of only phase modulation. The presence of both mixing schemes does not always

Classification task	Experimental Photonic ELM Mixing method			Software ELM
	PM [7]	**K**	**K + PM**	
Wine	97.5 %	99.3 %	98.9 %	96.1 %
Iris	93.9 %	98.3 %	99.6 %	96.0 %
Heart Disease	Not measured	83.2 %	79.1 %	81.0 %

Table 1: Accuracy measured on the benchmark tasks for the different configurations. PM: Phase Modulation; K: Kerr nonlinearity. The results described in column PM have been obtained in our previous experiment [7]. Experimental performances are compared with the performances measured on a similar software ELM (25 hidden neurons and quadratic nonlinearity).

produce an improvement in the results. This could be due to the nature of the tasks and to the decrease of the signal-to-noise ratio caused by the second phase modulator. The Kerr-based photonic ELM outperforms the software ELM: we believe that this is because the nonlinear mixing introduced by Kerr nonlinearity is more complex, and thus more effective, than the one provided by our simple software ELM.

Conclusion

We introduced and tested a novel scheme for a fiber-based Extreme Learning Machine (ELM). Information is encoded in the intensities of frequency comb lines and mixed by Kerr nonlinear effects arising from the propagation in a standard telecom single-mode fiber. This new scheme outperforms the previously proposed one [7], as well as a software ELM with a similar dimension but a quadratic nonlinearity. The advantage of this all-optical scheme (compared to the optoelectronic one based on phase modulation) will be further studied and explored in future work.

Acknowledgment

The authors acknowledge financial support from the European Commission grant 860360-POSTDIGITAL.

References

[1] M. M. Waldrop, “More than Moore,” Nature, vol. 530, issue 7589, pp. 144-148, 2016.
[2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature vol. 521, pp. 436–444, 2015.
[3] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: theory and applications,” Neurocomputing, vol. 70, issues 1-3, pp. 489–501, 2006.
[4] D. Pierangeli, G. Marcucci, and C. Conti, “Photonic extreme learning machine by free-space optical propagation,” Photonics Res., vol. 9, issue 8, pp. 1446–1454, 2021.
[5] U. Teğin, M. Yıldırım, İ. Oğuz, et al., “Scalable optical learning operator,” Nat. Comput. Sci., vol. 1, pp. 542–549, 2021.
[6] S. Ortín, M. Soriano, L. Pesquera, et al., “A Unified Framework for Reservoir Computing and Extreme Learning Machines based on a Single Time-delayed Neuron,” Sci Rep, vol. 5, no. 14945, 2015.
[7] A. Lupo, L. Butschek, and S. Massar, "Photonic extreme learning machine based on frequency multiplexing," Opt. Express, vol. 29, pp. 28257-28276, 2021
[8] R. A. Fisher., “Iris Data Set,” UCI Machine Learning Repository [Online], 1988. Available: https://archive.ics.uci.edu/ml/datasets/iris
[9] M. Forina et al., “Wine Data Set,” UCI Machine Learning Repository [Online], 1991. Available: https://archive.ics.uci.edu/ml/datasets/wine
[10] R. Detrano et al., “Heart Disease Data Set,” UCI Machine Learning Repository [Online], 1988. Available: https://archive.ics.uci.edu/ml/datasets/heart+disease