Research Article | Vol. 31, No. 12 / 5 Jun 2023 / Optics Express 20068
Optics EXPRESS

High-performance end-to-end deep learning IM/DD link using optics-informed neural networks

IOANNIS ROUMPOS,[1,2,*] LORENZO DE MARINIS,[3] MANOS KIRTAS,[4] NIKOLAOS PASSALIS,[4] ANASTASIOS TEFAS,[4] GIAMPIERO CONTESTABILE,[3] NIKOS PLEROS,[2,3] MILTIADIS MORALIS-PEGIOS,[2,3] AND KONSTANTINOS VYRSOKINOS[1,3]

[1]*Department of Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[2]*Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[3]*Center for Interdisciplinary Research & Innovation, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[4]*Scuola Superiore Sant'Anna, 56124, Pisa, Italy*
[*]*ioroumpo@auth.gr*

Abstract: In this paper, we introduce optics-informed Neural Networks and demonstrate experimentally how they can improve performance of End-to-End deep learning models for IM/DD optical transmission links. Optics-informed or optics-inspired NNs are defined as the type of DL models that rely on linear and/or nonlinear building blocks whose mathematical description stems directly from the respective response of photonic devices, drawing their mathematical framework from neuromorphic photonic hardware developments and properly adapting their DL training algorithms. We investigate the application of an optics-inspired activation function that can be obtained by a semiconductor-based nonlinear optical module and is a variant of the logistic sigmoid, referred to as the Photonic Sigmoid, in End-to-End Deep Learning configurations for fiber communication links. Compared to state-of-the-art ReLU-based configurations used in End-to-End DL fiber link demonstrations, optics-informed models based on the Photonic Sigmoid show improved noise- and chromatic dispersion compensation properties in fiber-optic IM/DD links. An extensive simulation and experimental analysis revealed significant performance benefits for the Photonic Sigmoid NNs that can reach below BER HD FEC limit for fiber lengths up to 42 km, at an effective bit transmission rate of 48 Gb/s.

1. Introduction

The recent resurgence of optical computing, fueled by both architectural and hardware breakthroughs [1], in conjunction with the growing success of Machine Learning (ML)-based approaches across the optical network and communication domain [2–4], are setting the stage for a new wave of converged ML-optics. In this framework, the combination of Deep Learning (DL) techniques with physical or artificial implementations of non-linear optical phenomena [5] are showing an enhanced performance [6]. In particular, autoencoder Neural Networks (NNs) have been gathering significant attention as a lower complexity, DL-based alternative in fiber communication link dispersion compensation [7–10]. Previous demonstrations have already showcased their credentials to enhance the reach and data rate of Intensity Modulation/Direct Detection (IM/DD) links compared to conventional equalization techniques [11]. All these DL models rely on the use of Rectified Linear Unit (ReLU) and ReLU-variants as their nonlinear activation functions, proceeding along the typical and successful paradigm of digital Artificial NNs (ANN).

#487209 https://doi.org/10.1364/OE.487209
Journal © 2023 Received 7 Feb 2023; revised 21 Apr 2023; accepted 5 May 2023; published 31 May 2023

Optics EXPRESS

More recently, however, the important progress reported by Deep Learning training models in adapting over the rapidly advancing neuromorphic photonic hardware [12–15] is shaping a new DL framework that has a strong potential to outperform digital ANN layouts in certain applications [16]. This DL framework is aligned to the idiosyncrasy of the underlying neuromorphic hardware, adapting to requirements and specifications not typically encountered in digital ANNs, i.e. analog noise [17], limited weight, input and output value range [1,18,19] and, last but not least, non-typical nonlinear activation functions [5,20]. The combined use of properly adapted DL training algorithms and analog photonic NN hardware can be referred to as an optics-informed or optics-inspired NN platform, which is governed by mathematical descriptions that stem from the physical laws of the photonic hardware. These have already shown certain performance benefits when applied in typical DL benchmarks [16,20] but have been never utilized for problems relating to optical communication and/or optical processing applications. Taking, however, into account that optical nonlinearities have been employed successfully in the past for signal restoration and regeneration purposes [21,22], the use of optics-informed nonlinear activation functions in DL engines that target optics-related problems could eventually strengthen the compensation of unwanted signal degradation phenomena.

In this paper, we propose and experimentally validate the performance of an optics-informed NN in an End-to-End DL-based IM/DD link, utilizing a non-linear activation function that is provided as a sigmoid-variant activation by a semiconductor-based optical module [5]. This nonlinear function, commonly referred to as "Photonic Sigmoid", follows a logistic sigmoid shape with a steep transition, with its physical implementation enabling also the parametrization of the photonic sigmoid slope. Through an extensive simulation analysis, we optimize the slope of the deployed photonic sigmoid activation function and benchmark its performance versus ReLU-based configurations that are commonly employed in state-of-the-art End-to-End DL systems. The simulation results along with the experimentally acquired data for transmission links up to 42 km at an effective rate of 48 Gb/s reveal improved performance both in noise- and dispersion-limited channels compared to respective ReLU-based demonstrations. A significant 3.5× transmission reach extension over state-of-the-art ReLU-based End-to-End DL schemes is experimentally validated for a Bit Error Rate (BER) reference of 10^{-6}, indicating the advantages of the optics-informed NN when utilized in a typical optics-dominated application field.

2. End-to-end communication link simulation model

We begin our analysis by developing the simulation model of an End-to-End communication channel as proposed in [8]. The model emulates an auto-encoder based IM/DD system, that demarcates from conventional PAM-M modulation formats by encoding the information in analog-time series [8], while also encapsulating the impairments of a communication link, such as nonlinearities arising from the square-law detection and fiber chromatic dispersion. The system is implemented as a deep fully connected feedforward ANN that mitigates the nonlinearities through the optimization of the transmitter-channel-receiver in a single End-to-End process. Figure 1 depicts the complete End-to-End ANN model comprised of three main parts: the transmitter, the nonlinear channel, and the receiver.

The Transmitter (Tx) NN can be viewed as the first 4 layers of the autoencoder, which processes the data in blocks. Each block, carrying $log_2(M)$ bits, is encoded in a one-hot vector of size M, and fed to the input layer of the Tx NN. The Tx NN consists of an input layer with M neurons, two hidden layers of 2 M neurons and an output layer with n neurons. Each block of $log_2(M)$ bits is transformed into a vector of n samples by this NN. In order to trigger the memory effect of the fiber nonlinear dispersion, during the training phase N = 11 consecutive blocks are fed into the channel per input forward propagation. Finally, the resulting blocks are concatenated by the serializer forming a waveform of N·n samples.

Optics EXPRESS

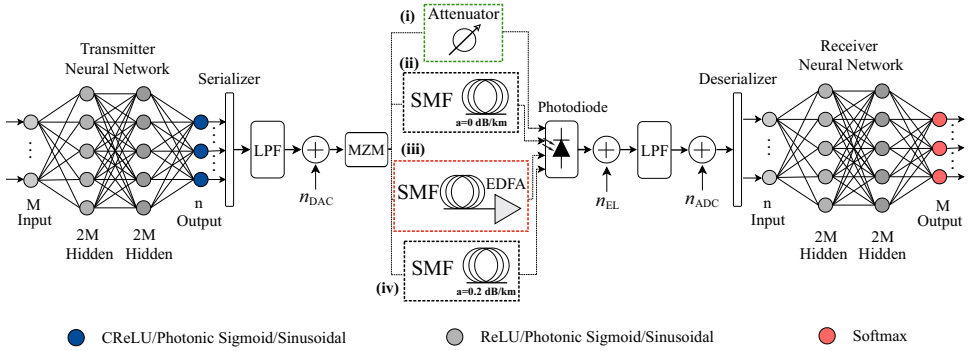

Fig. 1. Simulation model of the End-to-End IM/DD optical link comprising the ANN Tx and Rx, and the communication channel. Four different scenarios were evaluated (i) the *AWGN channel*, (ii) the *Dispersive channe*l where only chromatic dispersion is considered, (iii) the *Dispersive channel with constant AWGN*, where a constant power at the received is considered, and (iv) the *Dispersive AWGN channel*, that emulates a typical channel.

The output of the Tx NN is used to drive a Lithium Niobate ($LiNbO_3$) Mach-Zehnder Modulator (MZM). To account for the limited bandwidth of the opto-electronic components, the Tx NN output is passed through a Low-Pass brick wall Filter (LPF) with a 3-dB bandwidth of 36 GHz. Subsequently, zero mean Additive White Gaussian Noise (AWGN) (n_{DAC}) is added to the waveform, emulating the quantization noise of the employed Digital-to-Analog Converter (DAC), and correlated to its Effective Number of Bits (ENOB), through [23] :

$$\sigma_q = \sqrt{3P \cdot 10^{-(6.02 \cdot ENOB + 1.76)/10}} \tag{1}$$

where P corresponds to the average power of the optical signal. Finally, a sinusoidal function model $f_{MZM}(x) = \sin(x)$ imprints the electrical signal originating from the DAC on an optical Continues Wave (CW) light, emulating the transfer function of an MZM.

The signal emerging at the MZM output is injected into the communication channel that emulates optical fiber chromatic dispersion into the frequency domain through [24]

$$D(z, \omega) = e^{\left(j\frac{\beta_2}{2}\omega^2 z\right)} \tag{2}$$

where β_2 is the chromatic dispersion coefficient, ω is the angular frequency of the light wave and z corresponds to the fiber length. In order to quantify the effects of chromatic dispersion and channel noise when both of them are simultaneously present but also as stand-alone signal degradation factors, we consider four different communication channels for our analysis: (i) an *AWGN channel*, where a simple channel without fiber transmission, and as such zero chromatic dispersion, and variable losses was considered and emulated with an optical attenuator, (ii) a *Dispersive channel with zero AWGN,* where the channel comprises a Single Mode Fiber (SMF) of varying length, with the attenuation factor set to $a = 0\ dB/km$, in order to study the individual contribution of fiber chromatic dispersion, (iii) a *Dispersive channel with constant AWGN*, which corresponds to an automatic gain amplified transmission link, with variable SMF length and an attenuation coefficient of $a = 0.2\ dB/km$, where a constant output power level at the Rx front end is maintained. This channel is simulated through an automated gain Erbium Doped Fiber Amplifier (EDFA) with an Amplified Spontaneous Emission (ASE) noise negligible compared to the optical channel noise. (iv) A *Dispersive AWGN channel,* where a regular transmission link with varying SMF length is considered, with the attenuation coefficient set to $a = 0.2\ dB/km$ [25]. After propagation through the SMF, the model applies the photodetection square-law and introduces

Optics EXPRESS

AWGN n_{EL}, correlated to the photodiode (PD) input power and attributed to the following noise sources: Transimpedance Amplifier (TIA) thermal noise, shot noise, amplification noise and laser Relative Intensity Noise (RIN). It should be noted that $n_{EL,}$ is correlated with optical attenuation (α) introduced from the transmission link and correlated with the back-to-back experimentally measured noise (σ_{ref}) and the SMF length (ℓ) through:

$$\sigma_{n_{EL}} = \frac{\sigma_{ref}}{\sqrt{e^{(-\alpha\ell)}}} \tag{3}$$

Finally, the signal passes through an LPF with a 3-dB bandwidth of 36 GHz and quantization noise (n_{ADC}) is added, which emulates the effect of the Analog-to-Digital Converter (ADC).

At the ingress of the Receiver (Rx) NN, a deserializer separates the waveform in N = 11 blocks of n samples, isolating only the middle block for processing. The central block is launched into the Rx NN, with its dimensions mirroring in reverse the Tx NN. The final Rx NN layer exploits the softmax activation function, leading in a probability vector with a size of M, i.e. identical to the size of the initial One-hot vector. Finally, the BER of the system is calculated, and used as the main metric for the End-to-End communication channel performance.

Table 1 summarizes the parameters used for the simulation of the End-to-End optical communication channel. It should be noted that the sampling rate in the simulations was set at 320 GSa/s to provide an adequate resolution for the signal representation, capture in detail any spectral broadening effects [26] and correlate the 80 GSa/s rate of the experimentally available DAC with the simulation model.

Table 1. End-to-end simulation parameters

Simulation Parameters	
M	64
n	40
Sampling Rate	320 GSa/s
Symbol Rate	8 GSym/s
Information Rate	6 bits/sym
Effective Bit Rate	48 Gb/s
LPF bandwidth / type	36 GHz / Brick Wall
ENOB (DAC & ADC)	5
Fiber Dispersion coefficient (D)	16.2 ps/nm·km

During the optimization of the channel, the network was trained with a supervised learning style, evaluating the autoencoder ability to reconstruct random-generated bit sequences. The autoencoder was trained exploiting an ADAM optimizer, with variable learning rate, over a cross-entropy loss function. The models were trained for 1000 epochs on a set of $2.53{\cdot}10^4$ randomly generated blocks and validated on a set of $1.54{\cdot}10^4$ blocks, which have been generated with a different seed from that of the training data to avoid overfitting due to NN-related pattern recognition. [27] Towards benchmarking the effect of the activation function on the final achieved End-to-End performance, the simulation of each different channel was performed for three different activation functions configurations. In the first study, the previously demonstrated End-to-End paradigm was followed [8], with the Rectified Linear Unit (ReLU) applied in all layers except the Tx's NN final layer, where the Clipping ReLU (CReLU) was employed in order to clip the values in range of (0; pi/4), corresponding to the linear range of the MZM. The two optics-inspired activation functions that were additionally benchmarked are (*1*) a variant of the Photonic Sigmoid function reported in [5] and derived from semiconductor optical non-linearities and (*2*) a variant of the electro-photonic sinusoidal function clipped in the $y \in [0, 1]$,

Optics EXPRESS

corresponding to the electro-optic response of an MZM device [20]. In the photonic-sigmoid configuration the Tx's NN output was limited to the range [*0; π/2*], that given the incorporation of the MZM transfer function in the communication model resulted in no additional performance degradation. Figure 2 depicts a schematic presentation of the three different activation functions with $x \in [-1, 2]$ along with their mathematical expressions.

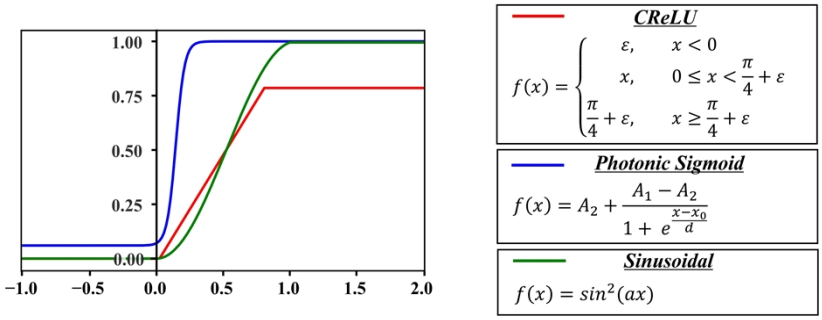

Fig. 2. CReLU, Photonic Sigmoid, Sinusoidal function plots along with their mathematical expressions.

To quantify the improved performance of an End-to-End DL channel embedding an optics-inspired activation function, we begin our analysis by quantifying the performance of the *Dispersive AWGN channel*, when the Tx and Rx NNs employ the photonic sigmoid in their hidden layers and the fiber length takes values in the {10-42} km range. The Photonic Sigmoid function can be described by the general expression of the logistic sigmoid function:

$$f(x) = A_2 + \frac{A_1 - A_2}{1 + \mathrm{e}^{(x - x_0)/d}} \tag{4}$$

with $A_1, A_{2,}, d$ derived by the experimental realization in [5] and set to $A_1 = 0.06$, $A_2 = 1.005$, $d = 0.033$. The parameter d relates to the slope of the logistic sigmoid function and can be tuned in the experimental realization through fine control of certain optical power levels [5], allowing for different variants of the photonic sigmoid to be studied as the activation function in the End-to-End DL network. Figure 3 (a) illustrates the achievable range of photonic sigmoid transfer functions within the slope tunability supported by its experimental implementation, while Fig. 3 (b) depicts the simulation results for the End-to-End *Dispersive AWGN* communication channel. The analysis reveals that the performance of the End-to-End channel, in terms of achieved BER, improves monotonically from $d = 0.023$ to $d = 0.073$, according to BER values at 35 km, and then plateaus and slowly decreases as d continues to increase up to $d = 0.113$. Based on this observation, throughout the rest of our study we employ the photonic sigmoid variant with $d = 0.073$.

Figure 4 (a)-(d) illustrates the acquired simulation results for the four communication channels. Specifically, Fig. 4 (a) depicts the obtained BER in the case of an *AWGN channel*, when the average optical power at the photodiode decreases from 3.5 to -9 dBm, with an attenuation factor that corresponds to the expected values when the signal propagates over a regular short reach link. Each label over the data points, up to -4.5 dBm, correlates the optical power reaching the Receiver PD with the length in km of the SSMF short reach link. As can be observed, the photonic sigmoid outperforms the ReLU/CReLU across the whole range of injected PD power, revealing the enhanced noise robustness that the steep profile can offer to NN autoencoders. Figure 4 (b) presents the simulation results for the *Dispersive channel with zero AWGN*, showing that the photonic sigmoid again significantly outperforms the ReLU/CReLU. Finally, Fig. 4 (c) and (d) depict the results for the *Dispersive Channel with constant AWGN* and the regular short reach

Optics EXPRESS

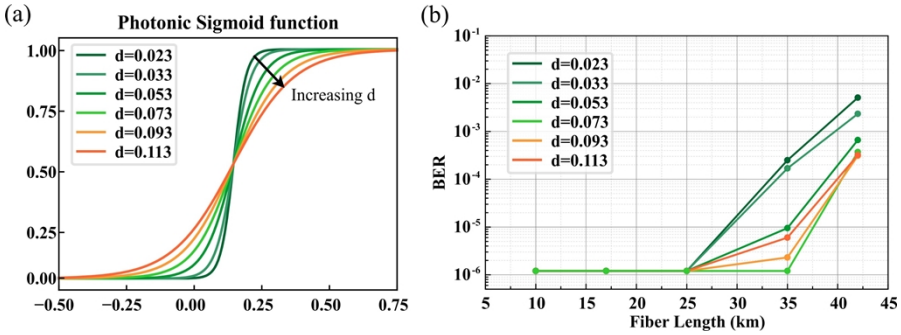

Fig. 3. (a) Photonic sigmoid function for different d values (b) Achieved BER for different d parameter values of the Photonic Sigmoid function tested on the *Dispersive AWGN channel*.

channel, respectively, revealing that the previously identified benefits of the photonic sigmoid extend to more complex channels that include both noise and chromatic dispersion. In the case of *Dispersive AWGN channel* six extra simulated fiber lengths are included for each activation function, in order to provide a more detailed insight into their response. Finally, it should be noted that the square-law of detection at the photodiode translates the chromatic dispersion and the noise of the channel in nonlinear impairments, with Fig. 4 results effectively showcasing that the photonic sigmoid offers significant benefits compared to the ReLU/CReLU and sinusoidal functions in non linear compensation tasks.

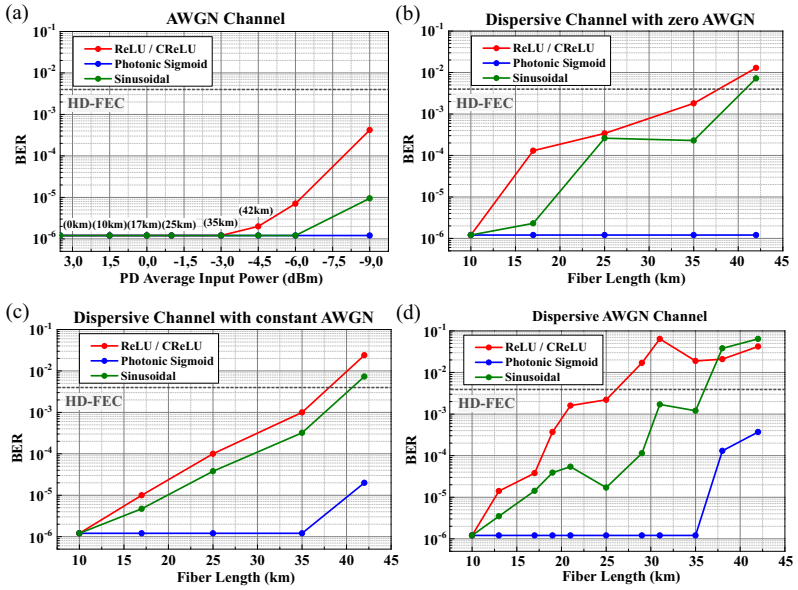

Fig. 4. Simulation results for (a) AWGN channel, (b) Dispersive channel with zero AWGN, (c) Dispersive channel with constant AWGN and (d) Dispersive AWGN channel including six extra simulated fiber lengths. In each case the performance of the examined activation functions (ReLU/CReLU, Photonic Sigmoid and Sinusoidal) is benchmarked.

Optics EXPRESS

3. Experimental evaluation and analysis of Dispersive and AWGN optical channels

Proceeding with the experimental evaluation of the end-to-end autoencoder we developed the experimental setup presented in Fig. 5, which allows the assessment of the different activation configurations for the three communication channels described in Section 2. The *Dispersive channel with zero AWGN* was excluded from the experimental study, as it would require zero noise levels that is practically unfeasible, while we focused on the ReLU/CReLU and the photonic sigmoid activation functions, as the highest performing functions from the "traditional" and optics-inspired paradigms. During the experimental testing, a randomly generated sequence of $5{\cdot}10^4$ elements was inserted in the Tx NN, while the appropriate trained model was applied. The output samples generated at 320 GSa/s were subsequently serialized and filtered in a 36 GHz bandwidth LPF. The LPF served a two-fold purpose: (i) Simulated the limited electro-optic

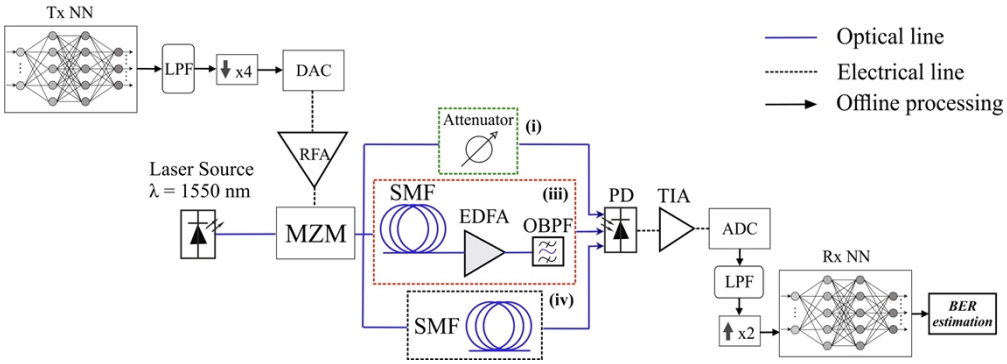

Fig. 5. Schematic of the experimental Setup employed for the system evaluation indicates the (i) *AWGN channel*, (ii) the *Dispersive channel with constant AWGN* and (iii) *Dispersive AWGN channel*.

bandwidth of the setup during training. (ii) Allowed preserving the spectral information of the input 320 GSa/s data stream, even when using the experimentally available 80 GSa/s DAC, as most of the spectral power density of the waveform is concentrated below 36 GHz. The resulting signal, after amplification with a Radio Frequency (RF) amplifier, was employed to drive an MZM acting on an incoming CW laser signal at λ=1550 nm, resulting in a data stream with an effective data rate of 48 Gb/s. In accordance with the studied simulation channels, we constructed their experimental analogous i.e., (i) *AWGN channel*, that allows isolating the optical channel and photodiode noise, while also containing an optical variable attenuator that adjust the input power at the photodiode. (iii) The *Dispersive Channel with constant AWGN*, employing a single mode fiber (SMF) spool of varying length, with attenuation and chromatic dispersion parameters of α=0.2 dB/km and D = 16.2 ps/nm·km, respectively. The optical fiber is followed by an automated gain EDFA, that maintains at its output the same optical power for different fiber lengths and an optical band pass filter (OBPF) with a 3-dB bandwidth of 1 nm for ASE noise rejection. (iv) The *Dispersive AWGN channel* that comprises a varying SMF spool. In all cases the optical signal is then converted to an electrical signal through a 70 GHz bandwidth photodiode, amplified in a TIA, digitized by a 160 GSa/s ADC and finally filtered in a digital brick wall type filter with 36 GHz bandwidth. The resulting recovered waveform is upsampled by factor of 2 to match the simulation model sampling rate of 320 GSa/s. The final waveform is deserialized and fed into the Rx NN. Prior final BER estimation, in order to account for the inevitable variations of the model and the experimental setup, the End-to-End NN is retrained in each case with a sequence of $2{\cdot}10^6$ symbols. Finally, the achieved BER of the system is estimated by comparing the final Rx NN layer with the corresponding training label, while it should be noted that the finite memory

Optics EXPRESS

of the DAC set a maximum number of transmittable samples and as such bound the minimum achievable BER value to $1.2{\cdot}10^{-6}$ with confidence interval of 95%. [28]

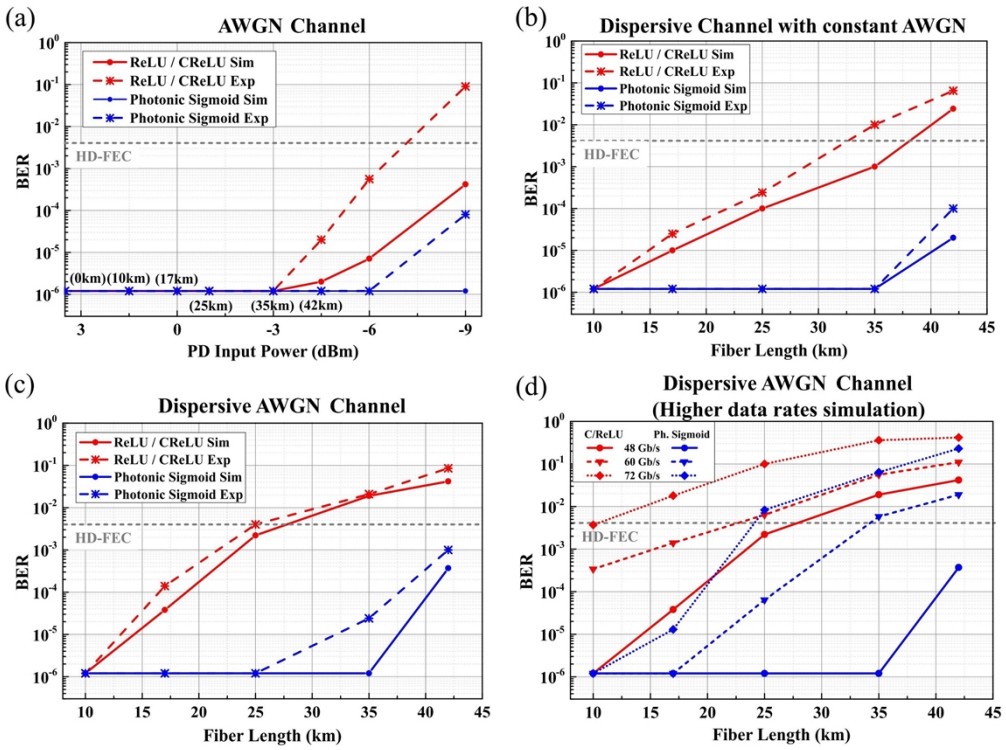

Fig. 6. Simulation and experimental BER values versus fiber testing length for the (a) AWGN channel, (b) the Dispersive channel with constant AWGN, (c) the Dispersive AWGN channel and (d) the simulation of Dispersive AWGN channel for higher information rates.

Figure 6 presents the experimentally obtained BER values for the three different communication channels and the two activation functions configurations plotted alongside the simulation estimations. In Fig. 6 (a), the results for the *AWGN channel* reveal that the Photonic Sigmoid achieves the minimum BER value at as low as -6 dBm of input optical power at the photodiode, outperforming the ReLU/CReLU, which can only reach this value for input power of -3 dBm. Figure 6 (b) illustrates the experimental acquired BER values for the *Dispersive channel with constant AWGN*, showcasing the Photonic Sigmoid achieves BER values well below the Forward Error Correction (FEC) limit ($4{\cdot}10^{-3}$) for fiber lengths up to 42 km while the corresponding value for ReLU/CReLU configuration barely surpasses 17 km. In the last scenario, where both chromatic dispersion and AWGN increase with fiber length, the Photonic Sigmoid is still able to provide a transmission system with BER values below the FEC limit for every testing fiber length, outperforming also in this case the ReLU/CReLU. In summary, the experimental analysis validated in all cases our theoretical projections, showcasing an excellent agreement with the simulation results and highlighting the clear benefits originating from the employment of the Photonic Sigmoid activation function in end-to-end trained optical communication channels. Finally, Fig. 6 (d) depicts the simulated BER performance of the two end-to-end autoencoder configurations, when the sampling rate is increased to 400 GSa/s and 480 GSa/s, resulting in aggregate information rates of 60 Gb/s and 72 Gb/s, respectively. The optics informed NN autoencoder significantly outperforms the ReLU across all tested rates and fiber lengths, allowing

Optics EXPRESS

for lower than HD-FEC at up to 25 km and 60 Gb/s, showcasing that using higher-end equipment can allow for even significant performance benefits.

Delving deeper into the NN performance in order to provide insights into the mechanism of operation that allows the photonic sigmoid to outperform the ReLU/CReLU, we plot in Fig. 7 (a) and (d) the Tx NN Output time traces in juxtaposition with the acquired traces before the Rx NN input, at 48 Gb/s and 35 km links, for both cases. Figure 7 (b) and (c) present 128 samples of the Tx and Rx 2nd hidden layer outputs, when ReLU is applied, while Fig. 7 (e) and (f) show the corresponding results for the Photonic Sigmoid. The four graphs reveal that the use of the Photonic Sigmoid forces the NN to operate in a more binary-like domain, as most of the samples are grouped in two discrete levels, while ReLU exhibits an analog-like response with samples amplitude taking more arbitrary values. The steep decision threshold of the Photonic Sigmoid assists the NN to converge better at the targeted values and as such allows for higher performance. This can be eventually also explained by the respective behavior of a steep logistic sigmoid function provided by semiconductor-based regenerative modules when applied in noise- and dispersion-dominated fiber communication channels, where their steep response and their strong power equalization properties facilitate the compensation of corresponding signal degradation effects. [21,29]

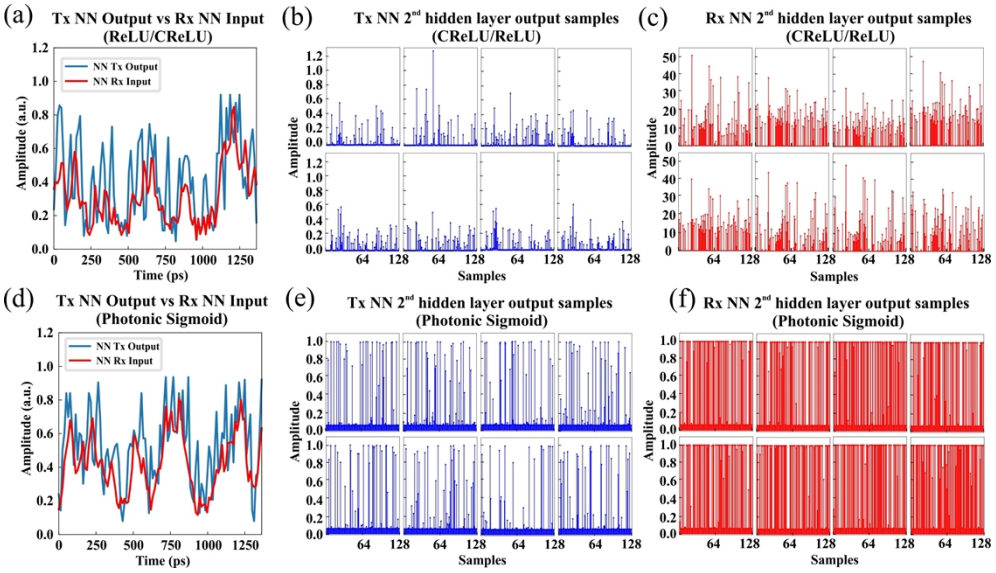

Fig. 7. (a) Acquired waveforms at 48 Gb/s of Tx NN output and Rx NN input for the ReLU/CReLU configuration at 35 km fiber length. Output values of the 128 nodes for 8 different samples at the (b) 2nd hidden layer of Tx NN and (b) at the 2nd hidden layer of Rx NN for the ReLU activation function. (d) Acquired waveforms of 48 Gb/s signal at the Tx NN output and Rx NN input for the Photonic Sigmoid at 35 km fiber length. Output values of the 128 nodes for 8 different samples at the (e) 2nd hidden layer of Tx NN and (f) at the 2nd hidden layer of Rx NN for the Photonic Sigmoid activation function.

4. Discussion

In order to unravel the underlying properties of optics informed NNs allowing them to outperform conventional implementations in chromatic dispersion compensation tasks, we benchmarked the performance of ReLu- and PhotonicSigmoid- based NNs, as function approximators of chromatic dispersion in a 10 km dispersive SSMF link.

Optics EXPRESS

Figure 8 (a) illustrates the employed 3-layer NN architecture, with the input, hidden and output layer comprising 2, N and 2 neurons, respectively. The input dataset was constructed using pairs of 2 samples of an input Gaussian pulse spanning in a time window of {75,75} ps where $P_{in} = e^{(-t^2/(T_0^2))}$, with $T_0 = 10$ ps, as illustrated at the inset of Fig. 8 (a). The output pulse after propagation through a 10 km SSMF was analytically calculated using the well-defined pulse-broadening equation [24]:

$$T_L(z) = T_0\sqrt{1 + \left(\frac{z}{L_D}\right)^2} \tag{5}$$

where z is the propagating distance set to 10 km, T_0 the initial duration of the pulse and $L_D = T_0^2/|\beta_2|$ the dispersion length that can be calculated through the dispersion coefficient of a typical SSMF fiber [25] at 1550 nm, resulting in D = 16.2 ps/nm·km and β_2= - 20.7·10^3 fs^2/m. The two NN configurations, employing the ReLu and Photonic Sigmoid activation function, respectively, were trained using an ADAM optimizer for 3000 epochs with a constant learning rate of 10^{-4}, with N being the number of neurons of the hidden layer ranging from {2-16}. Figure 8 (b) put in juxtaposition the achieved mean square error (MSE) for both configurations, revealing that the Photonic Sigmoid can achieve significantly lower MSE across all hidden layer variations, an effect more pronounced when the number of neurons in the hidden layer is lower than 16. Finally, Fig. 8 (c) offer a pictorial representation of the enhanced approximation capabilities of optics-informed NN in modelling chromatic dispersion, through overlapping the targeted output pulse versus the outputs of 8-neuron Relu- and Photonic Sigmoid based NNs. As can be clearly observed, the significant differentiation between the offered non-linearity of ReLU and underlying non-linear chromatic dispersion phenomenon results in lower performing NN deployments. Given that the main target of the end-to-end NN autoencoder is counteract the channel's degradation effects, a task equivalent with approximating the system inverse transfer function, it is evident that the higher function approximation properties of optics informed NNs, allow higher performance in optical communication applications.

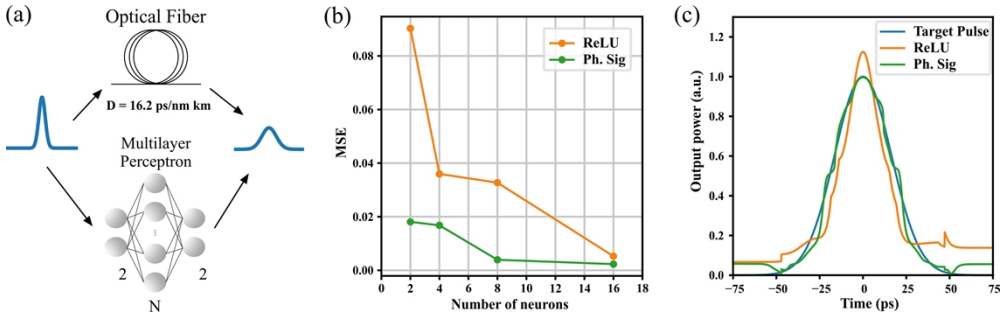

Fig. 8. (a) Scheme of the 3-layer perceptron that approximates the broadening of a gaussian pulse caused by the chromatic dispersion of a 10km optical fiber. (b) MSE values for the two different activation functions for different number of neurons in the middle layer of the perceptron. (c) Graph of the output pulses from a multilayer perceptron with 8 neurons at the hidden layer, plotted alongside the target pulse.

5. Conclusion

In this work we incorporate an optics-informed NN in an End-to-end Deep Learning network and demonstrate its application in an optical IM/DD fiber transmission link, leveraging an

Optics EXPRESS

optics-informed activation function for compensating the fiber chromatic dispersion and square-law detection related non-linearities. Our simulation and experimental analysis benchmarks the significant performance benefits of the Photonic Sigmoid trained NN versus the broadly implemented ReLU and ReLU-variants trained NN. The Photonic Sigmoid NNs can reach below BER HD-FEC limit for fiber lengths up to 42 km, at an effective bit transmission rate of 48 Gb/s. These outcomes validate that NNs, employing optics-derived activation functions, can act as a key enhancement factor in modern ML enhanced transmissions on IM/DD links that are governed by strong nonlinearities. Finally, the versatile properties of the Photonic Sigmoid offer enhanced thresholding properties ideally suited for deployment in short reach autoencoder-based NN, while its photonic implementation can be easily co-integrated alongside next generation transceivers.

Funding. Horizon 2020 Framework Programme (101092766, 871658).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. A. Tsakyridis, G. Giamougiannis, M. Moralis-Pegios, G. Mourgias-Alexandris, A. R. Totovic, M. Kirtas, N. Passalis, D. Lazovsky, A. Tefas, and N. Pleros, "Universal Linear Optics for Ultra-Fast Neuromorphic Silicon Photonics Towards Fj/MAC and TMAC/sec/mm2 Engines," IEEE J. Select. Topics Quantum Electron. **28**(6), 1–15 (2022).
2. F. Liao, S. Wei, and S. Zou, "Deep Learning Methods in Communication Systems: A Review," J. Phys.: Conf. Ser. **1617**(1), 012024 (2020).
3. J. W. Nevin, S. Nallaperuma, N. A. Shevchenko, X. Li, M. S. Faruk, and S. J. Savory, "Machine learning for optical fiber communication systems: An introduction and overview," APL Photonics **6**(12), 121101 (2021).
4. Y. Xie, Y. Wang, S. Kandeepan, and K. Wang, "Machine Learning Applications for Short Reach Optical Communication," Photonics **9**(1), 30 (2022).
5. G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, "An all-optical neuron with sigmoid activation function," Opt. Express **27**(7), 9620–9630 (2019).
6. Y. Miyashita, T. Kyono, K. Ikuta, Y. Kurokawa, and M. Nakamura, "Activation functions of artificial-neural-network-based nonlinear equalizers for optical nonlinearity compensation," *IEICE Communications Express advpub*, (2021).
7. B. Karanov, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end optimized transmission over dispersive intensity-modulated channels using bidirectional recurrent neural networks," Opt. Express **27**(14), 19650–19663 (2019).
8. B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-End Deep Learning of Optical Fiber Communications," J. Lightwave Technol. **36**(20), 4843–4855 (2018).
9. M. Li, D. Wang, Q. Cui, Z. Zhang, L. Deng, and M. Zhang, "End-to-end Learning for Optical Fiber Communication with Data-driven Channel Model," in *2020 Opto-Electronics and Communications Conference (OECC)* (2020), pp. 1–3.
10. Q. Zhang, Z. Wang, S. Duan, B. Cao, Y. Wu, J. Chen, H. Zhang, and M. Wang, "An Improved End-to-End Autoencoder Based on Reinforcement Learning by Using Decision Tree for Optical Transceivers," Micromachines **13**(1), 31 (2021).
11. B. Karanov, M. Chagnon, V. Aref, F. Ferreira, D. Lavery, P. Bayvel, and L. Schmalen, "Experimental investigation of deep learning for digital signal processing in short reach optical fiber communications," arXiv, arXiv.2005.08790 (2020).
12. B. J. Shastri, A. N. Tait, T. F. de Lima, W. H. P. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," Nat. Photonics **15**(2), 102–114 (2021).
13. G. Mourgias-Alexandris, M. Moralis-Pegios, A. Tsakyridis, S. Simos, G. Dabos, A. Totovic, N. Passalis, M. Kirtas, T. Rutirawut, F. Y. Gardes, A. Tefas, and N. Pleros, "Noise-resilient and high-speed deep learning with coherent silicon photonics," Nat. Commun. **13**(1), 5572 (2022).
14. L. el Srouji, A. Krishnan, R. Ravichandran, Y. Lee, M. On, X. Xiao, and S. J. ben Yoo, "Photonic and optoelectronic neuromorphic computing," APL Photonics **7**(5), 051101 (2022).
15. X. Guo, J. Xiang, Y. Zhang, and Y. Su, "Integrated Neuromorphic Photonics: Synapses, Neurons, and Neural Networks," Adv. Photonics Res. **2**(6), 2000212 (2021).
16. M. Kirtas, N. Passalis, G. Mourgias-Alexandris, G. Dabos, N. Pleros, and A. Tefas, "Robust Architecture-Agnostic and Noise Resilient Training of Photonic Deep Learning Models," IEEE Trans. Emerg. Top. Comput. Intell. **7**(1), 140–149 (2023).
17. M. Moralis-Pegios, G. Mourgias-Alexandris, A. Tsakyridis, G. Giamougiannis, A. Totovic, G. Dabos, N. Passalis, M. Kirtas, T. Rutirawut, F. Y. Gardes, A. Tefas, and N. Pleros, "Neuromorphic Silicon Photonics and Hardware-Aware Deep Learning for High-Speed Inference," J. Lightwave Technol. **40**(10), 3243–3254 (2022).

Optics EXPRESS

18. G. Mourgias-Alexandris, M. Moralis-Pegios, A. Tsakyridis, N. Passalis, M. Kirtas, A. Tefas, T. Rutirawut, F. Y. Gardes, and N. Pleros, "Channel response-aware photonic neural network accelerators for high-speed inference through bandwidth-limited optics," Opt. Express **30**(7), 10664–10671 (2022).
19. G. Giamougiannis, A. Tsakyridis, M. Moralis-Pegios, C. Pappas, M. Kirtas, N. Passalis, D. Lazovsky, A. Tefas, and N. Pleros, "Analog nanophotonic computing going practical: silicon photonic deep learning engines for tiled optical matrix multiplication with dynamic precision," (2023).
20. N. Passalis, G. Mourgias-Alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, "Training Deep Photonic Convolutional Neural Networks With Sinusoidal Activations," IEEE Trans. Emerg. Top. Comput. Intell. **5**(3), 384–393 (2021).
21. D. Apostolopoulos, K. Vyrsokinos, P. Zakynthinos, N. Pleros, and H. Avramopoulos, "An SOA-MZI NRZ Wavelength Conversion Scheme With Enhanced 2R Regeneration Characteristics," IEEE Photonics Technol. Lett. **21**(19), 1363–1365 (2009).
22. P. Zakynthinos, Ch. Kouloumentas, M. Bougioukos, P. Bakopoulos, E. Kehayas, A. Poustie, G. Maxwell, and H. Avramopoulos, "Multi-Format All-Optical Regeneration at 40 Gb/s Based on SOA-MZI," AIP Conf Proc **1288**(1), 93–96 (2010).
23. C. Pearson, "High-Speed, Analog-to-Digital Converter Basics", Texas Instruments, Dallas, TX, USA, App. Rep. SLAA510, (2011). Available online: http://www.ti.com/lit/an/slaa510/slaa510.pdf.
24. G. Agrawal, "*Fiber-Optic Communication Systems*", Fourth Edition, (Wiley, 2011).
25. Datasheet: "Corning SMF-28e+ Optical Fiber," Available online: https://www.corning.com/PI1463_07-14_English.
26. T. Xu, B. Karanov, and N. A. Shevchenko, *et al.*, "Digital nonlinearity compensation in high-capacity optical communication systems considering signal spectral broadening effect," Sci. Rep. **7**(1), 12986 (2017).
27. T. A. Eriksson, H. Bülow, and A. Leven, "Applying Neural Networks in Optical Communication Systems: Possible Pitfalls," IEEE Photonics Technol. Lett. **29**(23), 2091–2094 (2017).
28. Technical Article HFTA-010.0: Physical Layer Performance: Testing the Bit Error Ratio (BER), (2004). Available online: https://pdfserv.maximintegrated.com/en/an/3419.pdf.
29. M. Spyropoulou, N. Pleros, K. Vyrsokinos, D. Apostolopoulos, M. Bougioukos, D. Petrantonakis, A. Miliou, and H. Avramopoulos, "40 Gb/s NRZ Wavelength Conversion Using a Differentially-Biased SOA-MZI: Theory and Experiment," J. Lightwave Technol. **29**(10), 1489–1499 (2011).