Experimental Demonstration of a Dispersion Tolerant End-to-End Deep Learning-based IM-DD Transmission System

Mathieu Chagnon(1), Boris Karanov(1,2), and Laurent Schmalen(1)

(1) Nokia Bell Labs, Lorenzstraße 10, 70435 Stuttgart, Germany (mathieu.chagnon@nokia-bell-labs.com)
(2) University College London, London, WC1E 7JE, United Kingdom

Abstract *We experimentally demonstrate an IM-DD system relying on deep neural networks from transmitter to receiver delivering 42 Gb/s over 20 and 40 km at 1550 nm below 3.8×10^{-3}. The ANN is trained to tolerate deviations in dispersion by as much as ±170ps/nm.*

Introduction

The application of machine learning techniques in communication systems has attracted a lot of attention recently[1]. As an example, artificial neural networks (ANN) can be used to build probabilistic models of complex systems, serving in a variety of applications such as impairment characterization and monitoring as well as performance estimation and failure prediction[2,3]. ANNs are also known to be universal function approximators[4]. Fiber optic communication exhibiting chromatic dispersion (CD) and relying on direct detection (DD) are non-linear systems that can benefit from the robustness and flexibility of ANNs and are currently broadly investigated. One approach is to use ANNs for pre- and/or post-waveform distortion. For instance, ANNs are employed as receiver equalizers for PAM4 and PAM8 intensity modulation (IM) with direct detection (DD)[5,6] giving improvements in BER compared to using only a linear feed-forward equalizer. ANNs also allow to design communication systems by modelling the transmitter, receiver and channel as a whole and carrying out the optimization from end to end.

In this paper we build from our previous work[7] and present an experimental demonstration of an IM-DD transmission system solely relying on a multi-layered ANN from transmitter (Tx) waveform generation to receiver (Rx) symbol decision tolerating large variations in dispersion. We start by simulating a complete IM-DD transmission system delivering 42 Gb/s at 1550 nm targeting 20 or 40 km of SMF using the TensorFlow programming framework. The model includes all transceiver components. We train this end-to-end (E2E) ANN with varying link dispersions following a Gaussian distribution centered at the target distance, from which we obtain a set of Tx and Rx ANNs per target distance. We then use an optical transmission testbed with a fiber spool of 20 or 40 km to demonstrate the ANN-based IM-DD system. For each distance, the waveform generated at the Tx is the one obtained in simulation by the Tx ANN. By appropriately sweeping the dispersion with a tuneable dispersion module (TDM) and collecting real-time sampled photocurrents, we re-train the Rx ANN associated with this distance using experimental waveforms experiencing link dispersions following, once again, a Gaussian distribution. This contrasts with our previous work[7] where we re-trained the Rx ANN for every link dispersion, yielding a plethora of Tx-Rx ANN pairs that are each optimum for a very specific link condition but neither tolerant nor robust to any deviations from these conditions. We demonstrate BERs below 3.8×10^{-3} for both distances over a broad range of deviating dispersions. For 20 km, all BERs are below the threshold and as low as 10^{-5}, while results for 40 km are below threshold for large deviations spanning −150 to +90 ps/nm. Finally, we com-pare results against those from PAM2 (NRZ) and PAM4 giving the same bit rate and show that the ANN transceiver trained for 20 km outperforms NRZ and PAM4 after an equivalent distance of 16 and 27 km, respectively, while the ANN trained for 40 km outperforms both formats for any equivalent distance greater than 31 km. To the best of the authors' knowledge, this is the first experimental demonstration of a full end-to-end ANN-based IM-DD system delivering 42 Gb/s over 40 km at 1550 nm below 3.8×10^{-3} and tolerating large deviations in link dispersion.

Principles of End-to-End ANN IM-DD

A fully-connected ANN of L layers maps an input vector $\vec{x}$ to an output vector $\vec{y}$ through a series of L stages. The output of stage $k-1$ is the input to stage k, and the input-output relationship is

$$\vec{a}^{[k]} = f_{NL}^{[k]}\left(\mathbf{W}^{[k]}\vec{a}^{[k-1]} + \vec{b}^{[k]}\right), \qquad (1)$$

where $\mathbf{W}^{[k]}$ is an $n_k\times n_{k-1}$ matrix, $\vec{b}^{[k]}$ is a vector of length n_k, and $f_{NL}^{[k]}$ the non-linear function, also called *activation function*, of layer k and applied element-wise to $\vec{z}^{[k]} = \mathbf{W}^{[k]}\vec{a}^{[k-1]} + \vec{b}^{[k]}$. Vectors $\vec{a}^{[k-1]}$ and $\vec{a}^{[k]}$ are the input and output

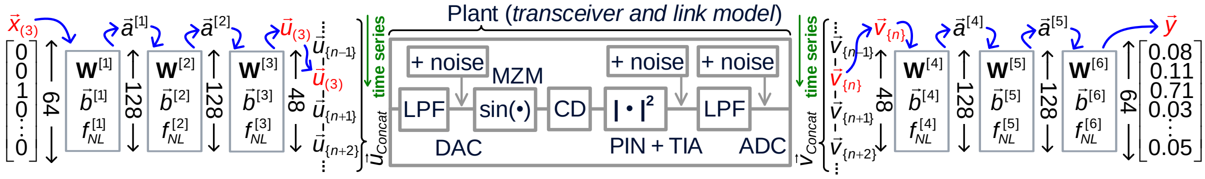

Fig. 1. Simulation test-bed for end-to-end deep-learning of IM-DD

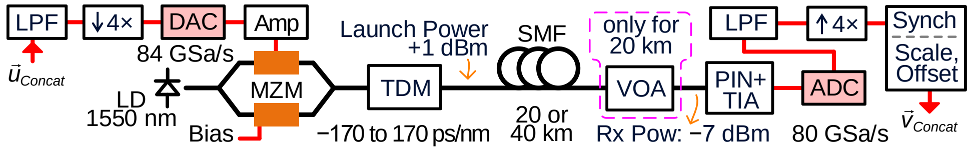

Fig. 2. Experimental testbed

for stage k of length n_{k-1} and n_k, respectively.

The model of our ANN-based end-to-end IM-DD system is depicted in Fig. 1. The system consists of a 3-layer ANN at the transmitter, a non-linear 'Plant', and a 3-layer ANN at the receiver. The 'Plant' is where we model all constituents of the IM-DD transmission system, namely, the digital to analogue converter (DAC), Mach-Zehnder Modulator (MZM), single mode fiber (SMF), photodetector (PIN), transimpedance amplifier (TIA), and analogue to digital converter (ADC). To limit the spectral extent of the signals generated by the DAC and digitized by the ADC, we add two brick-wall low-pass filters (LPF) of 32 GHz bandwidth at both ends of the Plant. We assume that the DAC and ADC have a finite ENOB of 6, that is mimicked by proper additive uniformly distributed noise. The PIN is simply modelled as square-law detector while the TIA is a source of additive white Gaussian noise (AWGN) representing thermal noise and setting the receiver noise floor.

Information encoding and decoding is done in the following way. A vector $\vec{x}_{(i)}$ of length 64 containing one '1' at index i and 63 '0' elsewhere, called a *one-hot vector*, passes through the 3-layer ANN of the Tx, yielding a vector $\vec{u}_{(i)}$ of length 48. There are 64 different one-hot vectors $\vec{x}_{(i)}$ and hence 64 different $\vec{u}_{(i)}$, each carrying 6 bit of information. In our simulation, the 48 samples of $\vec{u}_{(i)}$, forming a symbol, are generated at a rate of 336 GSample/s, or equivalently 7 GSymbol/s. N of these $\vec{u}_{(i)}$'s are randomly picked and concatenated to form a long vector of size 48N, called $\vec{u}_{Concat} = \ldots;\vec{u}_{\{n-1\}};\vec{u}_{\{n\}};\vec{u}_{\{n+1\}};\ldots$. In the example of Fig. 1, the n^{th} symbol transmitted $\vec{u}_{\{n\}}$ is $\vec{u}_{(3)}$. $\vec{u}_{Concat}$ passes through the Plant and exits as $\vec{v}_{Concat}$. The latter is decomposed into its constituting N vectors $\vec{v}$ of length 48, and each are propagated through the 3 layers of the receiver ANN.

In this work the activation functions $f_{NL}^{[k]}$ vary for the 6 layers. For layers k = 1, 2, 4 and 5, $f_{NL}^{[k]}$ is the rectified linear unit (ReLU) function, $f_{NL}^{[3]}$ is a windowed ReLU such that inputs $\vec{z}^{[3]}$ greater that $\pi/4$ are zeroed, and the last nonlinear function $f_{NL}^{[6]}$ is a *softmax* activation function.

For each final vector $\vec{y}$, we decide on the most probable transmitted one-hot vector $\vec{x}_{(i)}$ from the index of $\vec{y}$ having the largest value. The 6-bit label associated with $\vec{x}_{(i)}$ are Gray-coded with respect to the index. Note that this does not imply the entire transmission is Gray-coded. From the corresponding 6-bit label of the chosen $\vec{x}_{(i)}$'s, we compute the system's BER.

Simulation and Experimental Results

In the following, we present BER results from both simulation and experiment, for target distances around 20 and 40 km. The experimental test-bed built to validate this system is depicted in Fig. 2. A laser operating at 1550 nm is externally modulated by an MZM driven by $\vec{u}_{Concat}$ that is brick-wall low-pass filtered to 32 GHz. In the experiment, we 4-fold downsample the filtered $\vec{u}_{Concat}$, where each sample in the latter is now generated at a rate of 84 GSa/s rather than 336 GSa/s. There is no loss of information with downsampling thanks to the preceding LPF. The MZM is meticulously biased to match that of the simulation. The modulated optical signal first enters a TDM allowing to vary the dispersion from −170 to +170 ps/nm in steps in 10 ps/nm, before being launched at a power of +1 dBm in an SMF fiber spool of either 20 or 40 km. The optical link is unamplified. An AC-coupled PIN+TIA with a fixed input signal power of −7 dBm for both distances generates the photocurrent that is real-time sampled at 80 GSa/s by the ADC, itself of 32 GHz bandwidth. The digitized signal is upsampled back to 336 GSa/s. The only 'signal processing' applied on the received waveform $\vec{v}_{Concat}$ is rescaling and offset, such that $\vec{v}_{Concat}^{new} = \alpha\vec{v}_{Concat} + \beta$, which is necessary as the receiver has a different photo-responsivity than assumed in simulation and is AC-coupled. $\vec{v}_{Concat}^{new}$ is then sliced into blocks of 48 samples, each of which are forward-propagated through

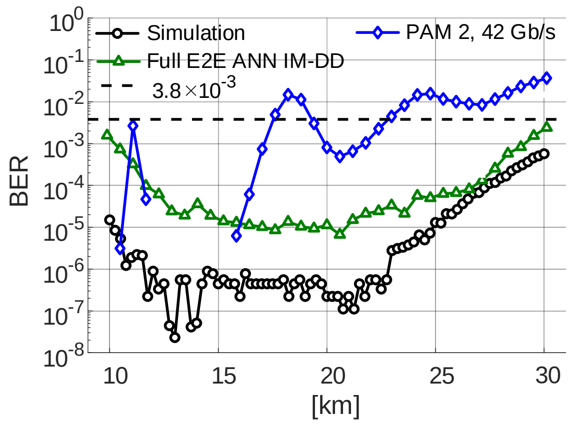

Fig. 3. Simulation and experimental results around 20 km of propagation.

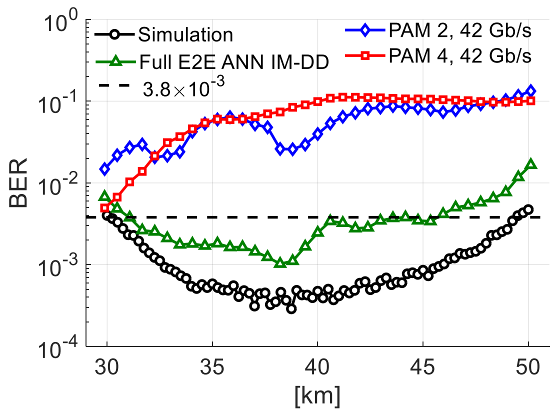

Fig. 4. Simulation and experimental results around 40 km of propagation

the receiver ANN followed by symbol decision, symbol-to-bit demapping and bit-error counting.

In this work, for each fiber spool, we re-train the three $\mathbf{W}$'s and $\vec{b}$'s of the receiver ANN, starting with their values obtained in simulation. We retrain with experimental waveforms for which the deviating dispersion, set by the TDM, follows a Gaussian distribution of 67 ps/nm standard deviation and spanning up to ±170 ps/nm. Using these re-trained Rx $\mathbf{W}$'s and $\vec{b}$'s, we test the performance for deviating dispersion ranging ±170 ps/nm and compute the BER.

Results are presented in Figs. 3 and 4 for varying dispersion with a 20 and 40 km fiber spool, respectively. The curves with round markers are simulation results. The curves with up-pointing triangle markers present experimental results of our full E2E ANN-based IM-DD system after re-training the Rx ANN. It is important to mention that these results are obtained from waveforms that are not found in the re-training set. Serving for comparison, we tested the performance of NRZ and PAM4 delivering the same bit rate and launched at the same power. For these, a T/2-spaced FFE equalizer of 13 taps was used at the receiver to help improve the performance and cover a temporal window at least as large as that covered by one $\vec{v}$. We observe from Fig. 3 that the ANN transceiver trained for 20 km outperforms NRZ and PAM4 after an equivalent distance of 16 and 27 km, respectively, assuming a dispersion coefficient of 16.8 ps/nm/km. In Fig. 4, we show that the ANN trained for 40 km outperforms both formats for any equivalent distance greater than 31 km.

One can observe that the performance of NRZ oscillates greatly in both figures. This is caused by the substantial variations of the power-fading profile for a 42 GBaud signal experiencing different link dispersions covering an extensive range of 340 ps/nm. On the other hand, the two ANN-based systems trained to tolerate dispersion variations do not present such strong dependence on the latter. Their performance follows a bathtub-like curve, as predicted from the simulations, and successfully demonstrate their robustness to deviations.

Conclusions

We presented an experimental demonstration of a C-band IM-DD transmission system delivering 42 Gb/s solely relying on a multi-layered ANN from transmitter waveform generation to receiver symbol decision. We trained a Tx and Rx ANN to tolerate large deviations in the link's dispersion, following a Gaussian distribution centered for either 20 or 40 km of SMF at 1550 nm. We built an IM-DD testbed as modelled and re-trained the receiver ANN once again with varying link dispersion. We obtained BERs for 20 km as low as 10^{-5} while those with a 40 km spool are below threshold for large deviations from −150 to +90 ps/nm. Both pairs of Tx-Rx ANNs give experimental BERs following the simulations and successfully demonstrate their robustness to variations in link dispersion.

Acknowledgement

EU Marie Skłodowska-Curie project COIN (676448/H2020-MSCA-ITN-2015).

References

[1] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 4, 2017

[2] F. Nadeem Khan, C. Lu, and A. Pak Tao Lau, "Machine Learning Methods for Optical Communication Systems," in Advanced Photon. Congr., paper SpW2F.3, 2017

[3] A. Saadaat Kashi, *et al.*, "Artificial Neural Networks for Fiber Nonlinear Noise Estimation," ACP 2017, Su1B.6

[4] K. Hornik, *et al.*, "Multilayer feedforward networks are universal approximators," Neural Netw., vol. 2, no. 5, 1989

[5] J. Estarán, *et al.*, Artificial Neural Networks for Linear and Non-Linear Impairment Mitigation in High-Baudrate IM/DD Systems," ECOC 2016, paper M.2.B.2

[6] S. Gaiarin, *et al.*, "High Speed PAM-8 Optical Interconnects with Digital Equalization based on Neural Network," ACP 2016, paper AS1C.1

[7] B. Karanov, *et al.*, "End-to-end Deep Learning of Optical Fiber Communications," http://arxiv.org/abs/1804.04097