End-to-End Learning in Optical Fiber Communications: Experimental Demonstration and Future Trends

Boris Karanov[(1,2,3)], Vinícius Oliari[(2)], Mathieu Chagnon[(3)], Gabriele Liga[(2)], Alex Alvarado[(2)], Vahid Aref[(3)], Domaniç Lavery[(1)], Polina Bayvel[(1)], and Laurent Schmalen[(4)]

[(1)] Optical Networks Group, Dept. of Electronic & Electrical Engineering, UCL, WC1E 7JE London, U.K.
[(2)] Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands b.p.karanov@tue.nl
[(3)] Nokia Bell Labs, 70435 Stuttgart, Germany
[(4)] Communications Engineering Lab, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Abstract *Fiber-optic auto-encoders are demonstrated on an intensity modulation/direct detection test-bed, outperforming state-of-the-art signal processing. Algorithms for end-to-end optimization using experimentally collected data are discussed. The end-to-end learning framework is extended for performing optimization of the symbol distribution in probabilistically-shaped coherent systems.*

Introduction

The design of communication transceivers as a single deep neural network (NN), optimized from end to end using deep learning[1]–[3] has attracted great interest in recent years[4],[5]. Such systems, referred to as *auto-encoders* (AE), are particularly suitable for communication over channels where the optimum transceiver is unknown or its implementation is computationally prohibitive. An important example is the application to optical fiber communications, for which advanced designs were developed, addressing the properties of the dispersive nonlinear channel[6]–[8]. It was shown that, applied to intensity modulation/direct detection (IM/DD) optical links, such systems can outperform, for a fixed processing memory of the algorithms, conventional pulse-amplitude modulation (PAM) transmission with receivers using state-of-the-art techniques such as nonlinear Volterra equalization or maximum likelihood sequence detection[7]–[11].

This paper complements the review of end-to-end deep learning-based optimization in optical fiber communications[6] by discussing the implementation and performance of the feedforward (FFNN) and sliding window bidirectional recurrent NN (SBRNN) transceiver on an IM/DD test-bed. The paper also explains the methods for performance enhancement via learning with measured data in both end-to-end and receiver-only mode. Experimental comparisons with state-of-the-art classical as well as deep learning-based DSP for IM/DD are discussed. Finally, to highlight future directions for end-to-end learning, an optimization of the input probability distribution in coherent fiber systems, modeled by the split-step Fourier method (SSFM), is performed.

AE on an actual transmission link

Within the AE framework, the channel is considered as a segment of the end-to-end computational graph representing the complete communication system[6]. Thus, the concept can be readily applied in scenarios where the channel model is known and differentiable. In this case, the transceiver is optimized via simulation transmission and applied "as is" to the real system, as shown in Fig. 1 i), presenting a viable perspective for low-cost systems since no training is required
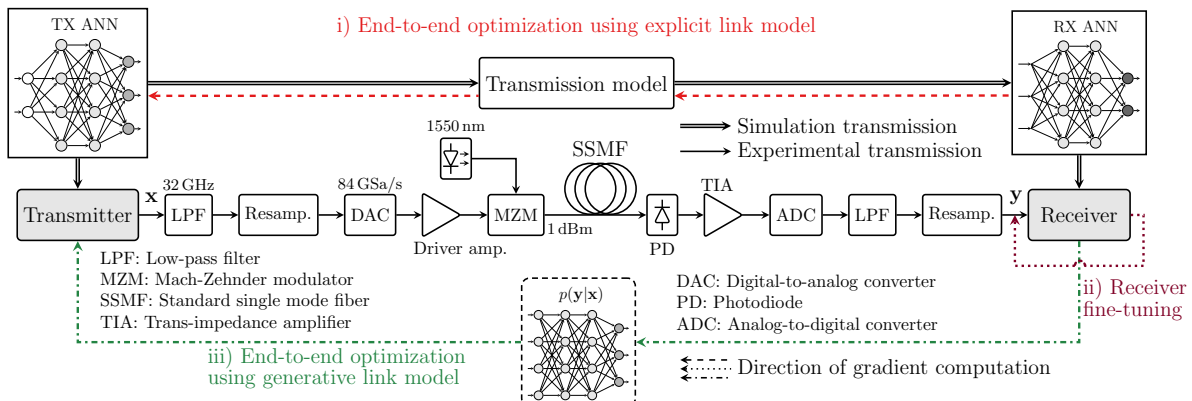

Fig. 1: Transmission test-bed and system optimization using i) numerical simulation of the link; ii) & iii) experimental traces.
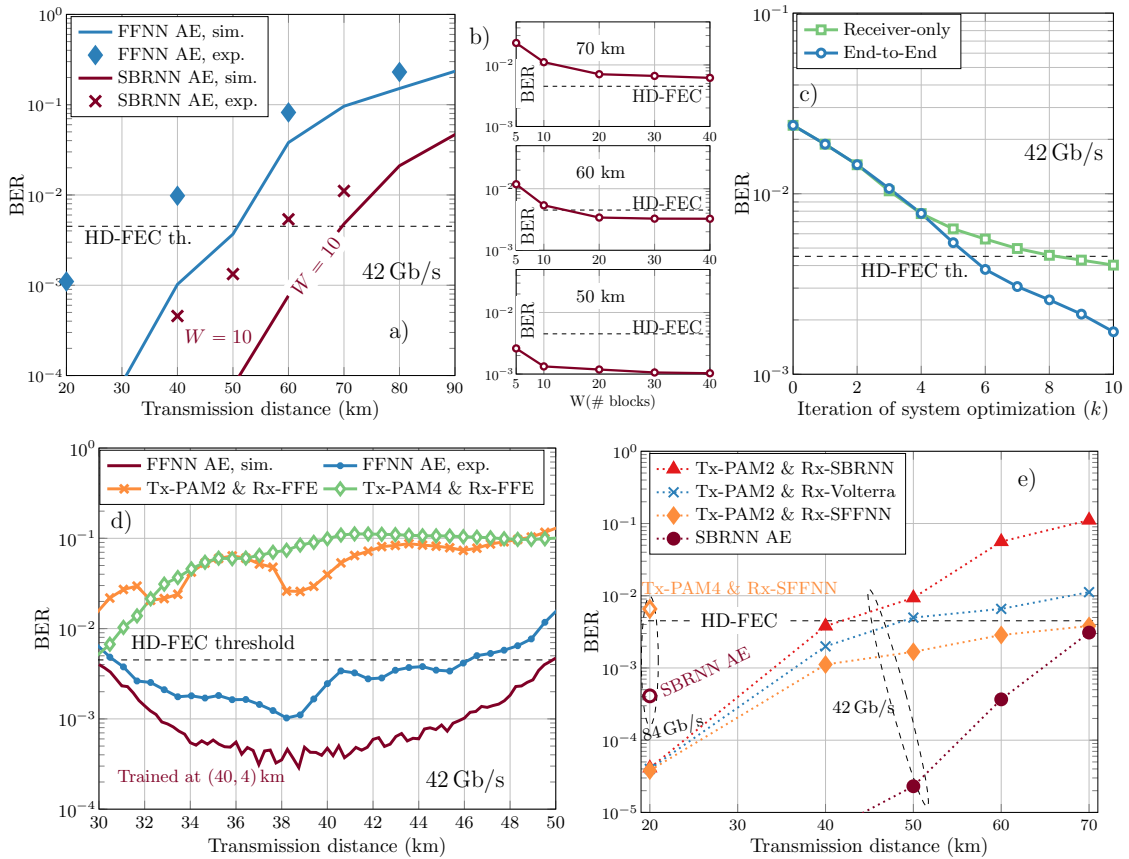
Fig. 2: a) BER vs. distance for FFNN and SBRNN AEs optimized in simulation-only; b) BER vs. processing window for 42 Gb/s SBRNN; c) BER vs. optimization step for FFNN AE optimized end-to-end using generative link model. d) BER vs. distance for the distance-agnostic FFNN transceiver. e) BER vs. distance for systems with receiver DSP optimized using experimental data.

after deployment. However, the performance of transceivers learned on a specific model assumption is often deteriorated when applied to an actual transmission link[3],[8],[9]. In such cases, the ANN parameters could be optimized to the specific link properties using experimental data. This is easier to implement at the receiver (Fig. 1 ii)), because the back-propagation[12],[13] for computing gradients extends only to the link output. The transmission link can be considered a *black box* for which only inputs and outputs are observed, precluding back-propagation to the transmitter parameters. To address this, optimization using a generative model of the optical link was developed and experimentally demonstrated[14]. It uses a generative adversarial network (GAN) for approximating the channel conditional distribution $p(\mathbf{y}|\mathbf{x})$. As shown in Fig. 1 iii), the obtained NN model is applied *in lieu* of the link during optimization, enabling back-propagation and computing transmitter gradients.

Experimental performance

FFNN and SBRNN AE were successfully demonstrated in optical IM/DD experiments and their performance was compared to PAM systems with

receivers using feedforward (FFE) and nonlinear Volterra equalizers as well as sliding window FFNN (SFFNN) and BRNN[8]–[10]. The test-bed schematic is shown in Fig. 1. The 32 GHz low-pass filtered transmitter output was applied to the DAC, with the obtained electrical waveform modulating a 1550 nm laser. After fiber propagation, the received waveform was direct-detected, real-time sampled and stored for receiver DSP.

In Fig. 2 a) the BER performance of 42 Gb/s FFNN and SBRNN AE was compared for systems optimized only in simulation[7],[8]. The SBRNN showed significantly superior performance, allowing transmission below the 6.7% HD-FEC[15] at 50 km distance. The system reach was enhanced to 60 km by adjusting the sliding window size to $W = 20$ (Fig. 2 b)). In agreement with simulation[6], diminishing gains for greater W were observed since system non-linearities and noise start to dominate when ISI is compensated. AE and PAM schemes were compared when the receiver DSP is optimized using experimental data. As seen in Fig. 2 e) this reduced the BER of the 42 Gb/s SBRNN AE below HD-FEC at 70 km and allowed 84 Gb/s at 20 km. Compared to
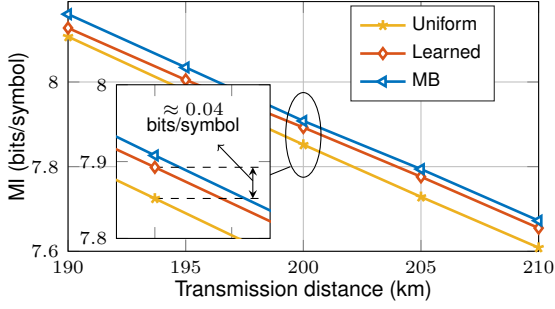
Fig. 3: Mutual information vs. distance for uniform QAM, Maxwell-Boltzmann (MB), learned distribution.

Fig. 4: MB and learned probability mass functions for 1024-QAM at the transmission distance of 200 km.

classical nonlinear Volterra equalization as well as state-of-the-art deep learning-based DSP for PAM, AE increased reach or enhanced data rate at shorter distances for fixed algorithm processing memory. The results from experiment, verifying a novel optimization method for transceivers *robust* to distance variation[8],[10] are shown in Fig. 2 d). The *distance-agnostic* AE achieved BERs below HD-FEC for wide range (> 15 km) of distances around the nominal 40 km, significantly outperforming PAM with FFE receivers, optimized *separately* for each length. Finally, the results from the experiment of end-to-end system optimization using a generative model of the link are shown in Fig. 2 c), observing a monotonically decreasing BER at each step. The algorithm outperformed receiver optimization on measured data, showing increasing gains as optimization progressed.

Extenson of the AE framework

The optical AE framework can be extended towards different systems and models. For example, it has been applied for geometrical shaping (GS) using simplified models for coherent systems[16]–[18]. In this section, the concept of an optical system as an end-to-end computational graph is used for learning the symbol distribution in probabilistically-shaped (PS) coherent links modeled by SSFM[19]. For such a model, multiple Fourier/inverse Fourier transform steps form segments of the graph, making optimization more challenging[20],[21]. PS optimization within AE was recently proposed for AWGN channels[22],[23]. For fiber communications, PS was highlighted as an effective method for reach and throughput increase as well as rate adaptation[24]. Nevertheless, finding the optimal distribution for the fiber channel is an open problem, especially for multi-dimensional modulation formats[25]. The scheme used for PS distribution optimization in this work applies the techniques proposed in[22] to fiber communications. A waveform up-sampled by a factor of 8 was considered to accurately capture
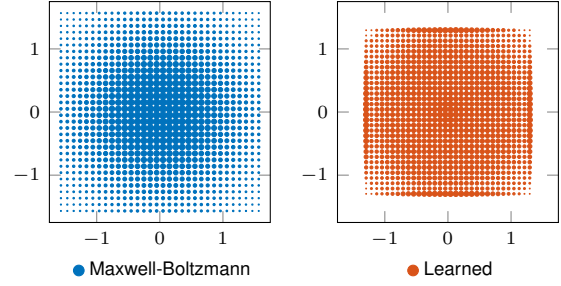
the nonlinear effects. The system consisted of a single span, single polarization 33 GBaud signal with dispersion, attenuation and nonlinear fiber coefficients of 16.3 ps/nm/km, 0.163 dB/km, and 1.2 1/W/km, respectively, emulating the link of[26]. At the transmitter, a single ANN layer was used to optimize the logits which yield the distribution probabilities. The layer dimensionality is that of the transmitted 1024-QAM constellation, kept fixed for this investigation. The signal is pulse-shaped by a 0.1 roll-off RRC filter. After matched filtering and sampling, chromatic dispersion compensation is applied at the receiver before the ANN. The computed loss was the categorical cross-entropy between the receiver outputs and the one-hot vectors representing the transmitted constellation, corrected by the constellation entropy. This loss can approximate the mutual information (MI)[22], which we used as a performance metric. The launch power was optimized together with the ANNs. Optimization was performed using the Adam algorithm[27] with meticulously adjusted learning rate and consisted of 4 batches of 2^{14} transmit symbol sequences. Fig. 3 shows the MI as a function of distance for the uniform, Maxwell-Boltzmann (MB), and learned distributions. We observe that the learned distribution outperforms the uniform by 0.04 bits/symbol at 200 km with performance close to MB. Interestingly, Fig. 4 shows that the learned distribution at 200 km significantly differs from MB. The presented results are the first steps of an ongoing investigation of end-to-end learning applied to PS for optical systems as well as joint PS and GS.

Conclusions

Different approaches for implementing auto-encoders in optical transmission links are discussed. In an IM/DD test-bed, reach increase or enhanced data rate are demonstrated as well as robustness to distance variations compared to classical DSP. Moreover, deep learning is applied to coherent systems for optimizing the symbol distribution in PS coherent optical systems.

Acknowledgements

References

[1] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention", in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2016, pp. 223–228.

[2] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer", *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[3] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep learning based communication over the air", *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, 2018.

[4] F. N. Khan, Q. Fan, C. Lu, and A. P. T. Lau, "An optical communication's perspective on machine learning and its applications", *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 493–516, 2019.

[5] S. Cammerer, F. Ait Aoudia, S. Dörner, M. Stark, J. Hoydis, and S. ten Brink, "Trainable communication systems: Concepts and prototype", *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5489–5503, 2020.

[6] B. Karanov, P. Bayvel, and L. Schmalen, "End-to-end learning in optical fiber communications: Concept and transceiver design", in *Proc. of European Conference on Optical Communication (ECOC)*, 2020, pp. 1–4.

[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*, vol. 27, no. 14, pp. 19 650–19 663, Jul. 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", *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018.

[9] 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", in *Proc. of IEEE International Workshop on Signal Processing Systems (SiPS)*, 2020, pp. 1–6.

[10] M. Chagnon, B. Karanov, and L. Schmalen, "Experimental demonstration of a dispersion tolerant end-to-end deep learning-based IM-DD transmission system", in *Proc. of European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.

[11] B. Karanov, G. Liga, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "Deep learning for communication over dispersive nonlinear channels: Performance and comparison with classical digital signal processing", in *Proc. of 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2019, pp. 192–199.

[12] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, pp. 533–536, 1986.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT press, 2016.

[14] B. Karanov, M. Chagnon, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "Concept and experimental demonstration of optical IM/DD end-to-end system optimization using a generative model", in *Proc. of Optical Fiber Communications Conference (OFC)*, 2020, pp. 1–3.

[15] Z. Wang, "Super-FEC codes for 40/100 Gbps networking", *IEEE Communications Letters*, vol. 16, no. 12, pp. 2056–2059, 2012.

[16] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities", in *Proc. of European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.

[17] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape", in *45th European Conference on Optical Communication (ECOC 2019)*, 2019, pp. 1–4.

[18] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information", in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.

[19] J. A. C. Weideman and B. M. Herbst, "Split-step methods for the solution of the nonlinear Schrodinger equation", *SIAM Journal on Numerical Analysis*, vol. 23, no. 3, pp. 485–507, 1986.

[20] T. Uhlemann, S. Cammerer, A. Span, S. Dörner, and S. ten Brink, "Deep-learning autoencoder for coherent and nonlinear optical communication", *arXiv:2006.15027*, 2020.

[21] S. Gaiarin, F. Da Ros, R. T. Jones, and D. Zibar, "End-to-end optimization of coherent optical communications over the split-step Fourier method guided by the nonlinear Fourier transform theory", *arXiv:2009.05324*, 2020.

[22] M. Stark, F. Ait Aoudia, and J. Hoydis, "Joint learning of geometric and probabilistic constellation shaping", in *IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[23] F. Ait Aoudia and J. Hoydis, "Joint learning of probabilistic and geometric shaping for coded modulation systems", *arXiv:2004.05062*, 2020.

[24] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration", *Journal of Lightwave Technology*, vol. 34, no. 7, pp. 1599–1609, 2016.

[25] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "On shaping gain in the nonlinear fiber-optic channel", in *2014 IEEE International Symposium on Information Theory*, 2014, pp. 2794–2798.

[26] E. Sillekens *et al.*, "A simple nonlinearity-tailored probabilistic shaping distribution for square QAM", in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, 2018, pp. 1–3.

[27] D. Kingma and J. Ba, "Adam: A method for stochastic optimization", in *Proc. of International Conference for Learning Representations*, 2015, pp. 1–15.