End-to-End Learning in Optical Fiber Communications: Concept and Transceiver Design

Boris Karanov(1,2), Polina Bayvel(1), and Laurent Schmalen(3)

(1) Optical Networks Group, Dept. of Electronic & Electrical Engineering, UCL, London, U.K.
(2) Signal Processing Systems Group, Eindhoven University of Technology, Eindhoven, Netherlands
(3) Communications Engineering Lab, Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract *We implement a complete fiber-optic communication system as an end-to-end computational graph using an artificial neural network (ANN)-based transceiver. We highlight transceivers implemented using feedforward or recurrent ANN, and illustrate their performance by an example.*

Introduction

Conventional communication systems consist of several digital signal processing blocks, each performing individual tasks at the transmitter and receiver, e.g. coding, modulation, pulse-shaping or equalization. However, there is a lack of optimal, computationally feasible algorithms for communication over dispersive nonlinear channels found in the optical fiber systems. Consequently, a transceiver with one or several sub-optimal modules may not achieve the optimal end-to-end system performance, limiting the achievable data rates and transmission reach. Designing such systems requires carefully chosen approximations. The combination of artificial neural networks (ANNs), known as universal function approximators[1], and deep learning[2] provides a framework for optimizing the system in a single end-to-end process – an idea first introduced for wireless communications[3]–[5]. The approach was quickly utilized also in optical fiber communications aiming at exploiting to a greater extent the potential for data transmission over nonlinear dispersive channels[6]–[14]. It consists in implementing the complete fiber-optic system as an end-to-end computational graph using ANN-based transmitter and receiver. The method avoids the modular design of conventional fiber-optic systems and enables a transceiver structure that can be optimized for a specific metric in a single deep learning process spanning from the transmitter input to the receiver output.

This paper highlights the approach of performing end-to-end deep learning in optical fiber communications. It discusses two transceiver designs tailored for communication over dispersive nonlinear channels using feedforward or bidirectional recurrent ANNs and compares their complexity. We examine the systems performance for short-reach optical fiber links based on intensity modulation/direct detection (IM/DD) – preferred in many data center, metro and access networks, and impaired by the dispersion-induced interference and nonlinear photo-detection[15].

Optical fiber system as an end-to-end graph

The framework of end-to-end system optimization via deep learning is based on interpreting the complete chain of transmitter, channel and receiver as a *computational graph*, shown in Fig. 1. This is achieved by using a differentiable link model and implementing the transceiver as a deep ANN to form the graph segments. Such a design allows us to jointly optimize the signal processing in an end-to-end process over the constraints imposed by the channel. In particular, the transmitter ANN *encodes* the sequence of random input messages from a finite alphabet $(\ldots m_t \ldots)$ into a sequence of symbols $(\ldots \mathbf{x}_t \ldots)$, where a symbol is a block of multiple *waveform* samples.

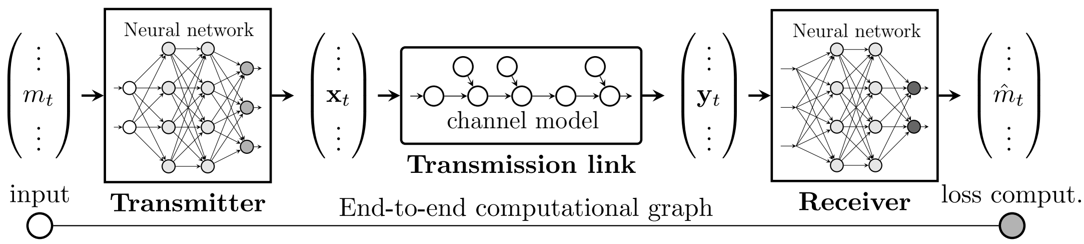

Fig. 1: Optical fiber communication system implemented as an end-to-end computational graph.

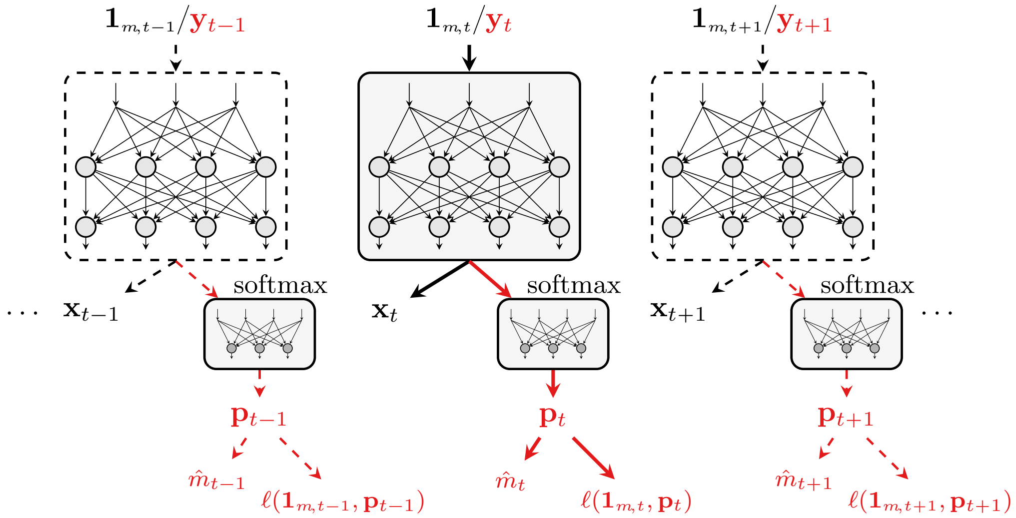

Fig. 2: Schematic of the symbol encoding (black arrows) and decoding (red arrows) by an FFNN-based auto-encoder.

The produced digital waveform is fed to the transmission link, acquiring noise as well as interference according to the channel model. The channel output is distorted waveform samples, forming the received symbols $(\ldots\mathbf{y}_t\ldots)$. These are *decoded* by the receiver ANN, obtaining the recovered messages $(\ldots\hat{m}_t\ldots)$. During optimization, a loss between the transmitted and received messages $\mathcal{L}=\sum_t \ell(m_t,\hat{m}_t)$ is computed. Utilizing the dependencies on the end-to-end computational graph, the back-propagation algorithm[16] is applied to obtain gradients of $\mathcal{L}$ with respect to the trainable transceiver ANN parameters. The transceiver is optimized via gradient descent[17],[18] aimed at minimising $\mathcal{L}$, e.g. the message (symbol) error rate. Such *auto-encoder* systems are suitable in communication scenarios where the optimum transmitter-receiver pair is unknown or computationally prohibitive.

Transceiver design and performance

Chromatic dispersion causes interference from both preceding and succeeding samples and renders the optical fiber a channel with memory[15],[19], necessitating the processing of data sequences. To tailor the transceiver for communication over such a channel, auto-encoders based on feedforward (FFNN)[6] as well as bidirectional recurrent (BRNN) ANNs[10] have been proposed. Figure 2 shows a schematic of transmitter and receiver designed as an FFNN. This network processes each input independently and the architecture can be used to encode/decode symbols without information from the previous or future encoding/decoding. At the transmitter, the message m_t, independently chosen from a set of M messages, is represented as a *one-hot vector* $\mathbf{1}_{m,t}\in\mathbb{R}^M$, for which the m-th element is 1 and all others are 0. It is applied to the FFNN, whose input dimension is M. After processing by multiple hidden layers, the output of the final layer, a vector with dimension n, represents the encoded block of samples (symbol) $\mathbf{x}_t$. The FFNN encoding can be denoted as $\mathbf{x}_t=f_{\text{E-FFNN}}\left(\mathbf{1}_{m,t}\right)$. At the receiver, the n-dimensional symbol $\mathbf{y}_t$ is decoded to a probability vector $\mathbf{p}_t\in\mathbb{R}^M$, i.e. $\mathbf{p}_t=f_{\text{D-FFNN}}\left(\mathbf{y}_t\right)$, using another multi-layer FFNN with input and *softmax*[2] output dimensions n and M, respectively. The vector $\mathbf{p}_t$ is utilised in two ways: during transceiver optimization, the end-to-end system loss is computed as $\mathcal{L}(\theta)=\frac{1}{|S|}\sum_{t\in S}\ell(\mathbf{1}_{m,t},\mathbf{p}_t)$, where θ is the set of transceiver parameters, S the set of transmitted messages and $\mathbf{p}_t\triangleq f_{\text{D-FFNN}}\left(\mathcal{H}\left\{f_{\text{E-FFNN}}\left(\mathbf{1}_{m,t}\right)\right\}\right)$ expresses the complete input-to-output FFNN auto-encoder mapping with $\mathcal{H}\{\cdot\}$ describing the channel. The output $\mathbf{p}_t$ is further used for calculating the symbol error rate as $\text{SER}=\frac{1}{|S|}\sum_{t\in S}\mathbb{1}\left\{m_t\neq\text{argmax}(\mathbf{p}_t)\right\}$. The FFNN auto-encoder is inherently unable to compensate for interference outside of the symbol block, which is treated as extra noise. As a consequence, the achievable performance in terms of compensated interference and hence transmission distance is limited by the block size. An alternative design based on BRNNs, utilizing information from both pre- and post-cursor symbols, is shown in Fig. 3. The input is concatenated with the previous/next encoded symbol to produce $\overrightarrow{\mathbf{x}}_t=f^{\text{fw}}_{\text{E-RNN}}\left(\left(\mathbf{1}^T_{m,t}\ \ \overrightarrow{\mathbf{x}}^T_{t-1}\right)^T\right)$ or $\overleftarrow{\mathbf{x}}_t=f^{\text{bw}}_{\text{E-RNN}}\left(\left(\mathbf{1}^T_{m,t}\ \ \overleftarrow{\mathbf{x}}^T_{t+1}\right)^T\right)$. These are merged via element-wise averaging into

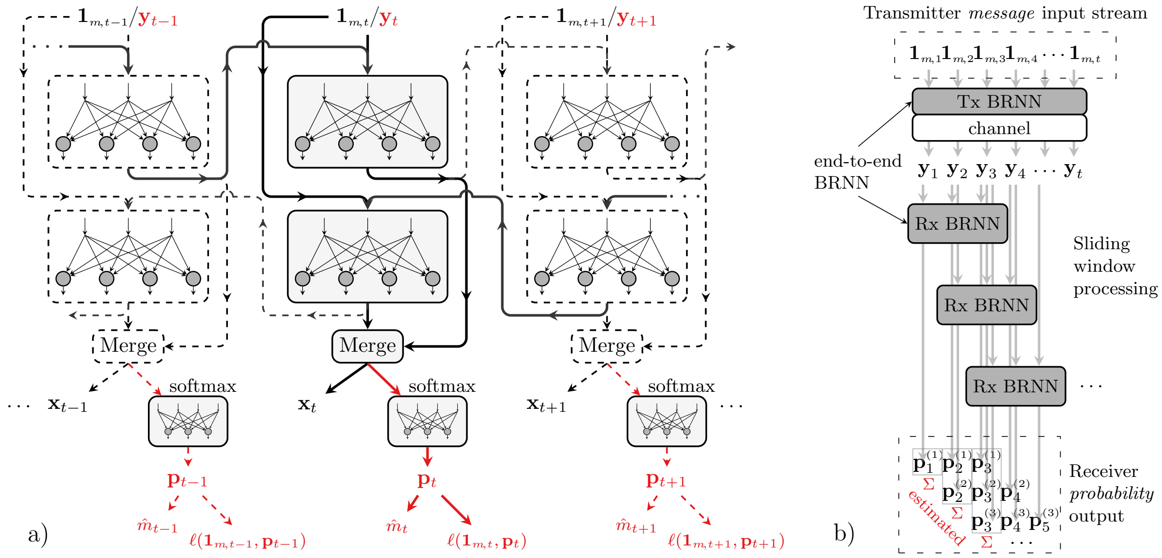

Fig. 3: a) Symbol encoding (black) and decoding (red) by a BRNN. b) Sliding window sequence estimation.

the n-dimensional encoded symbol $\mathbf{x}_t$. The RNNs in both directions have dimensionality $M+n$ (input) and n (output). For decoding, the received symbol $\mathbf{y}_t$ is concatenated with the preceding/succeeding receiver output, producing $\overrightarrow{\mathbf{h}}_t = f^{\text{fw}}_{\text{D-RNN}}\left(\begin{pmatrix}\mathbf{y}_t^T & \overrightarrow{\mathbf{h}}_{t-1}^T\end{pmatrix}^T\right)$ or $\overleftarrow{\mathbf{h}}_t = f^{\text{bw}}_{\text{D-RNN}}\left(\begin{pmatrix}\mathbf{y}_t^T & \overleftarrow{\mathbf{h}}_{t+1}^T\end{pmatrix}^T\right)$. These are merged into $\mathbf{h}_t$ via concatenation. The receiver RNNs have dimensionality $n+2M$ (input) and $2M$ (output). *Softmax* is applied to $\mathbf{h}_t$, obtaining the output probability $\mathbf{p}_t \in \mathbb{R}^M$, utilized for computing the system loss $\mathcal{L}(\theta) = \sum_t \ell(\mathbf{1}_{m,t}, \mathbf{p}_t)$, where $\mathbf{p}_t \triangleq f_{\text{D-BRNN}}\left(\mathcal{H}\left\{f_{\text{E-BRNN}}\left(\ldots \mathbf{1}_{m,t-1}, \mathbf{1}_{m,t}, \mathbf{1}_{m,t+1} \ldots\right)\right\}\right)$ is the BRNN auto-encoder function. The optimized transceiver is employed in a sliding window scheme[20] (SBRNN), shown in Fig. 3 b). The transmitter encodes the *full* stream of input messages, while at time t, the receiver BRNN decodes the *window* of W symbols $(\mathbf{y}_t, \ldots, \mathbf{y}_{t+W-1})$ to $(\mathbf{p}_t^{(t)}, \ldots, \mathbf{p}_{t+W-1}^{(t)})$. The final output probabilities for message decision and error counting are estimated as $\mathbf{p}_i = \sum_{k=0}^{\min(W,i)-1} [\min(W,i)]^{-1} \cdot \mathbf{p}_i^{(i-k)}$, i.e. the receiver scheme provides multiple estimates for the symbol at time t, which are combined.

The rate of the auto-encoders is $\rho = \log_2(M)/n$. Expanding the FFNN processing memory for fixed ρ translates in larger n and $M = 2^{\rho \cdot n}$, rapidly increasing the number of trainable parameters. In contrast, the BRNN receiver memory depends only on the window W, external to the end-to-end network architecture, whose number of parameters can be fixed. The auto-encoders are applied to optically un-amplified IM/DD links where deep

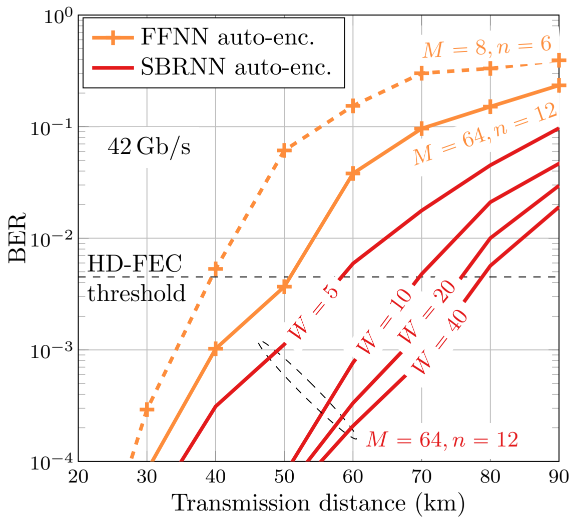

Fig. 4: BER versus distance for the FFNN and SBRNN auto-encoders applied to optical IM/DD communication.

learning is seen as a viable DSP for addressing the dispersion and square-law detection induced limitations[21]. The link model starts with 84 GSa/s DAC followed by a modulator, and finishes with ADC[6],[10]. The BER[11] of the systems at 42 Gb/s for different M, n and W is compared in Fig. 4. Adjusting W, the SBRNN allows transmission below the 6.7% HD-FEC[22] at distances beyond 70 km – yielding $>$ 20 km improvement over FFNN. Verified in experiments, the systems outperformed state-of-the-art DSP[6],[13],[14].

Conclusions

This paper reviews the methods for end-to-end optimized optical fiber transmission. It discusses the signal processing and complexity in auto-encoders based on feedforward and recurrent ANNs. The presented designs are general and can be applied to different systems and models.

Work under EU MSCA COIN (676448/H2020-MSCA-ITN-2015) & UK EPSRC TRANSNET (EP/R035342/1)

References

[1] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning", in *Proc. of European Conference on Optical Communication (ECOC)*, 2018, pp. 1–3.

[10] 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.

[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] 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.

[13] 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.

[14] B. Karanov, M. Chagnon, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "Optical fiber communication systems based on end-to-end deep learning", in *Proc. of IEEE Photonics Conference (IPC)*, 2020, pp. 1–2.

[15] G. Agrawal, *Fiber-Optic Communication ystems, 4th ed*. Wiley, 2010.

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

[17] H. B. Curry, "The method of steepest descent for non-linear minimization problems", *Quarterly of Applied Mathematics*, vol. 2, no. 3, pp. 258–261, 1944.

[18] S. Ruder, "An overview of gradient descent optimization algorithms", *arXiv:1609.04747*, 2016.

[19] E. Agrell, A. Alvarado, G. Durisi, and M. Karlsson, "Capacity of a nonlinear optical channel with finite memory", *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2862–2876, 2014.

[20] N. Farsad and A. Goldsmith, "Neural network detection of data sequences in communication systems", *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5663–5678, 2018.

[21] V. Houtsma, E. Chou, and D. van Veen, "92 and 50 gbps TDM-PON using neural network enabled receiver equalization specialized for PON", in *Proc. of Optical Fiber Communications Conference and Exhibition (OFC)*, 2019, pp. 1–3.

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