Neural Turbo Equalization to Mitigate Fiber Nonlinearity

Koike-Akino, T.; Wang, Y.; Millar, D.S.; Kojima, K.; Parsons, K.

TR2019-088 September 05, 2019

Abstract

We propose a turbo equalization scheme based on deep neural networks (DNN) to compensate for fiber nonlinearity. The turbo DNN equalizer can accelerate decoding convergence and achieve a significant gain of about 2 dB in nonlinear regimes.

European Conference on Optical Communication (ECOC)

NEURAL TURBO EQUALIZATION TO MITIGATE FIBER NONLINEARITY

Toshiaki Koike-Akino* , Ye Wang, David S. Millar, Keisuke Kojima, Kieran Parsons

Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA.
**E-mail: koike@merl.com*

Keywords: DEEP LEARNING, TURBO EQUALIZATION, DIGITAL SIGNAL PROCESSING, FIBER NONLINEARITY, HIGH-ORDER QAM

Abstract

We propose a turbo equalization scheme based on deep neural networks (DNN) to compensate for fiber nonlinearity. The turbo DNN equalizer can accelerate decoding convergence and achieve a significant gain of about 2 dB in nonlinear regimes.

1 Introduction

In fiber-optic communications, we encounter various linear/nonlinear impairments, such as laser linewidth, amplified spontaneous emission, chromatic dispersion, polarization mode dispersion, self-phase modulation, cross-phase modulation, four-wave mixing, and cross-polarization modulation. To realize high-speed, reliable, and long-reach optical communications, a number of nonlinear equalization methods to compensate for such distortion were investigated, e.g., turbo equalizer (TEQ) [1–3], Volterra series [4], and digital backpropagation (DBP) [5–8]. As an alternative to those equalization schemes, machine learning techniques have recently been envisioned to play a viable role in mitigating fiber nonlinearity [9], e.g., Gaussian mixture models [10], particle method [7], independent component analysis [11], hidden Markov models [12], support vector machines (SVM) [13] and shallow/deep neural networks (DNN) [14–25]. In particular, deep learning techniques have shown its high potential in nonlinear performance improvement, e.g., with end-to-end design [22–25].

Nonetheless, most work did not appropriately account for practical interaction with forward error correction (FEC) codes. For example, multi-class soft-max cross-entropy loss is often used to train DNN, which assumes nonbinary FEC codes in principle. For more practical bit-interleaved coded modulation (BICM) systems, it was found in [19] that binary sigmoid cross-entropy loss can improve accuracy and scalability to high-order quadrature-amplitude modulation (QAM) [19]. In this paper, we propose another DNN application to perform TEQ for nonlinear mitigation in the context of BICM with iterative demodulation (ID). Although DNN has already been popular in nonlinear compensation, our paper is the first attempt to adopt DNN for TEQ in the context of BICM-ID which takes soft-decision feedback from the FEC decoder to refine the DNN output for improved equalization accuracy. We make an analysis of the extrinsic information transfer (EXIT) of turbo DNN, and demonstrate that the proposed DNN paired with irregular LDPC codes used in DVB-S2 standards offers a significant performance gain of about 2 dB by accelerating the decoder convergence in nonlinear transmissions.

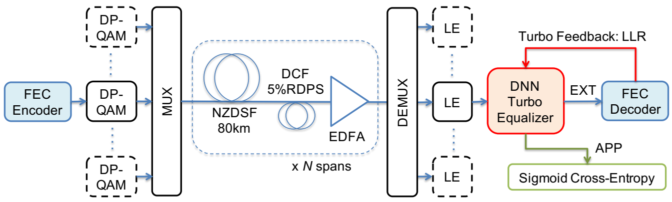
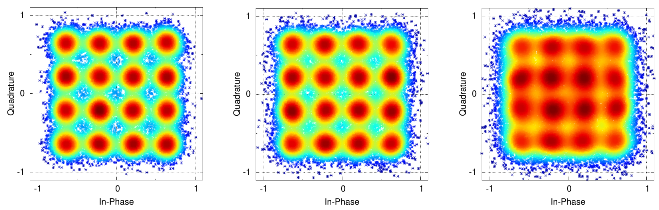


Fig. 1: Coherent optical communications with DNN-TEQ.



(a) −5 dBm Launch (b) −3 dBm Launch (c) −1 dBm Launch

Fig. 2: Residual nonlinear distortion of DP-16QAM constellation after LE for 16-span NZDSF DM links.

2 Deep Learning for Nonlinear Compensation

The optical communications system under consideration is depicted in Fig. 1. Multi-channel DP-QAM signals with wavelength multiplexing are sent over fiber plants towards coherent receivers. We consider N spans of dispersion managed (DM) links with 80 km non-zero dispersion-shifted fiber (NZDSF) at a residual dispersion per span (RDPS) of 5%. The span loss is compensated by Erbium-doped fiber amplifiers (EDFA). The receiver employs standard phase recovery and linear equalization (LE) to compensate for linear impairments such as chromatic dispersion. Due to fiber nonlinearity, residual distortion after LE will limit the achievable information rates.

Fig. 2 shows an example of residual distortion of DP-16QAM constellation after 31-tap least-squares LE for 16-span transmissions. We can see that the constellation is more seriously distorted with the increased launch power due to Kerr

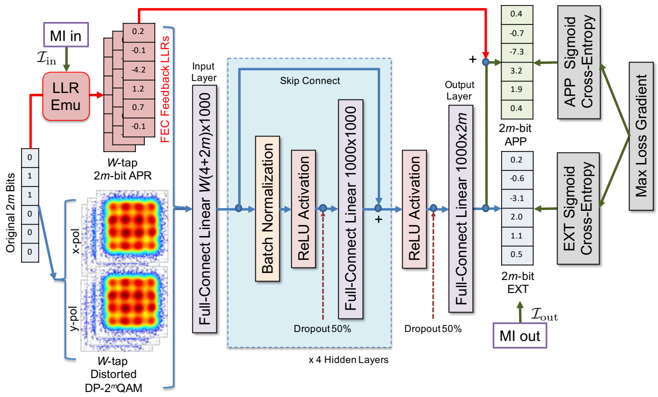


Fig. 3: DNN-TEQ architecture and min-max-loss training.

fiber nonlinearity. To compensate for the residual nonlinear distortion, we introduce deep learning-based TEQ, which exploits soft-decision feedback from FEC decoder as shown in Fig. 1.

Deep learning has been studied as a breakthrough technique in media processing research, where many-layer many-node neural architectures are trained with a large amount of data. Note that big data are readily available in high-speed optical communications, which can provide terabits of data in a second. The DNN is massively parallelizable in hardware, which is suited for future optical communications. In modern DNN, various techniques have been introduced, e.g., pre-training, mini-batch, rectified linear unit (ReLU), dropout, skip connection, inception, adaptive-momentum (Adam) stochastic gradient, adversarial, convolutional, and long short-term memory (LSTM) architectures, In this paper, we employ state-of-the-art DNN suited for BICM-ID to cope with fiber nonlinearity.

3 Turbo DNN Equalization: DNN-TEQ

Fig. 3 shows the architecture of our turbo DNN equalizer, which feeds distorted DP-QAM signals over consecutive $W = 3$-tap symbols to generate soft-decision log-likelihood ratio (LLR) values for FEC decoding. The major extension from conventional DNN lies in the input layer which takes *a priori* (APR) side information along with DP-QAM symbols. The APR side information comes from FEC decoder representing intermediate soft-decision LLRs in run time. For efficient DNN training, the APR values having mutual information of $\mathcal{I}_{\rm in}$ are synthetically generated via a Gaussian distribution following $\mathcal{N}((-1)^b\sigma^2/2, \sigma^2)$ where b is an original bit and $\sigma = J^{-1}(\mathcal{I}_{\rm in})$ with $J^{-1}(\cdot)$ being ten Brink's J-inverse function [26], instead of considering a particular FEC decoder feedback.

The last layer has two branches, i.e., *extrinsic* (EXT) output and *a posteriori* probability (APP) output, which uses a skip connection from the input layer to sum up EXT and APR at a target symbol. This residual network tries to train extrinsic message passing for TEQ realization. It was found that learning DNN model to minimize APP cross-entropy loss does not always minimize EXT cross-entropy loss accordingly, and vice versa. In order to keep both APP and EXT outputs reliable, we use a max-pooling layer following sigmoid cross-entropy loss.

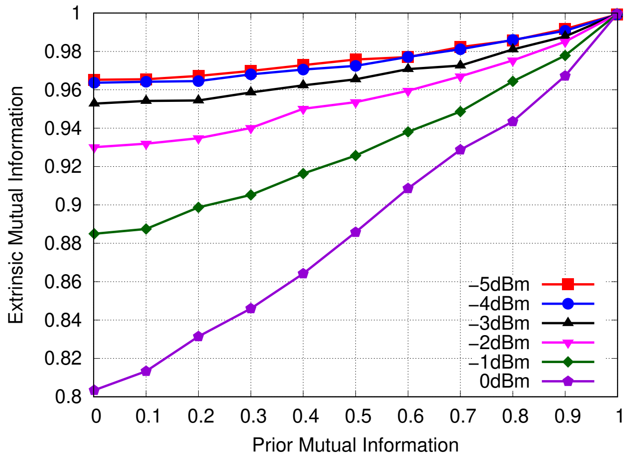


Fig. 4: EXIT chart of DNN-TEQ for DP-16QAM in 16-span DM links.

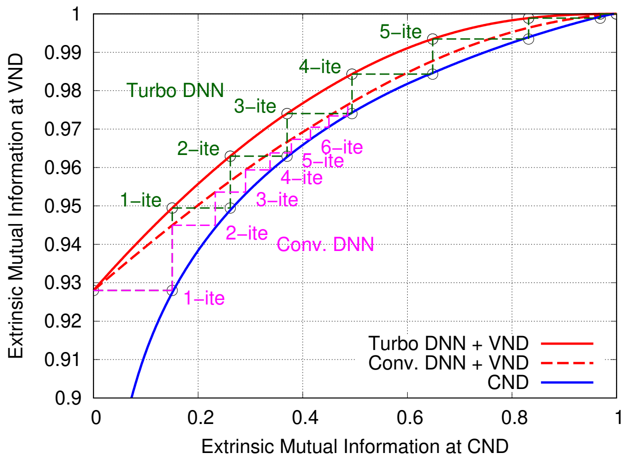


Fig. 5: Combined EXIT chart [26] of DNN-TEQ & LDPC decoder for DVB-S2 code rate $9/10$ (DP-16QAM in 16-span DM links at -2 dBm).

The DNN uses four hidden layers, each of which consists of batch normalization, ReLU activation, and a fully-connected linear layer with skip connections and 50% dropout for 1000 neuron nodes. The DNN is trained with Adam for a mini-batch size of 1000 symbols to minimize the worst sigmoid cross-entropy losses between APP and EXT outputs, using training datasets of approximately 5×10^5 symbols. An early stopping with a patience of 13 is carried out up to a maximum of 500 epochs. Note that sigmoid cross-entropy minimization is equivalent to maximizing the lower bound of the generalized mutual information (GMI), which is an important metric for soft-decision FEC performance.

4 Performance Results

We assume 3-channel DP-QAM transmission for 34 GBd baud rate and 37.4 GHz channel spacing, over NZDSF DM links with 5% RDPS, having a dispersion parameter of $D = 3.9$ ps/nm/km, a nonlinear factor of $\gamma = 1.6$ /W/km, and an attenuation of 0.2 dB/km. Span loss is compensated by ideal EDFA with all amplified spontaneous emission noise added just before the receiver assuming the noise figure of 5 dB. We used digital root-raised cosine filters with 10% rolloff at both transmitter and receiver.

Besides DNN equalizers, we compare various classical machine learning methods, such as linear discriminant analysis (LDA), naïve Bayes (NB), quadratic discriminant analysis (QDA), and SVM. For FEC codes, we consider variable-rate irregular LDPC codes of block length 64,800 bits, used in DVB-S2 standards. The LDPC codes have a different degree distribution for individual code rates. For instance at a code rate of $9/10$, the variable degree polynomial (node perspective) is given as $\lambda(x) = 0.1x^2 + 0.8x^3 + 0.1x^4$, whereas the check degree is 30. Although the degree distribution can be optimized jointly with DNN-TEQ as done analogously in [26], we leave it as the future work.

Fig. 4 shows the EXIT chart of DNN-TEQ given LLRs having a certain mutual information from the FEC decoder. It is clearly observed that the DNN outputs can be greatly improved by feeding in the FEC soft-decision. An almost linear slope towards $\mathcal{I}_{\mathrm{out}} = 1$ in EXIT curve is achieved, implying that cross-entropy loss is mitigated linearly with FEC feedback reliability. This steep slope in the EXIT curve of DNN-TEQ can eventually make a significant improvement in LDPC decoding performance, as shown in Fig. 5, where we present the decoding trajectory between the variable-node decoder (VND) and the check-node decoder (CND) in the LDPC decoder. Here, we use a combined EXIT chart [26] of DNN-TEQ and LDPC decoder, for DP-16QAM 16-span DM links at -2 dBm launch power and DVB-S2 LDPC codes with a code rate of $9/10$. As shown, the conventional DNN equalizer without FEC feedback requires a large number of decoder iterations to reach an error-free mutual information of $\mathcal{I}_{\mathrm{out}} = 1$. Whereas for DNN-TEQ, we can open up an EXIT tunnel between VND and CND, that leads to a considerable acceleration of the decoder convergence to reach error-free condition within only a few iterations.

Figs. 6, 7, and 8 show the Q factor versus launch power of DP-4QAM, DP-16QAM, and DP-64QAM, respectively, for 50, 16, and 8 spans. It is observed that DNN or LSTM can offer superior performance to classical learning methods. Note that LSTM had no gain over DNN because channel memory is limited in DM links and LE already shortens the memory. With the proposed turbo DNN architecture, we can further improve the performance by up-to 2.7 dB at the peak Q.

5 Conclusions

We extended DNN machine learning techniques to TEQ for improved nonlinear compensation in coherent fiber communications. Through EXIT chart analysis, we verified that the proposed DNN-TEQ offers decoder acceleration by feeding intermediate soft-decision LLR from the LDPC decoder. It was found that our DNN-TEQ improves Q factor through the turbo iteration by a gain of about 2 dB in nonlinear regimes. To the best of authors' knowledge, this is the first paper investigating TEQ based on DNN for fiber nonlinearity mitigation.

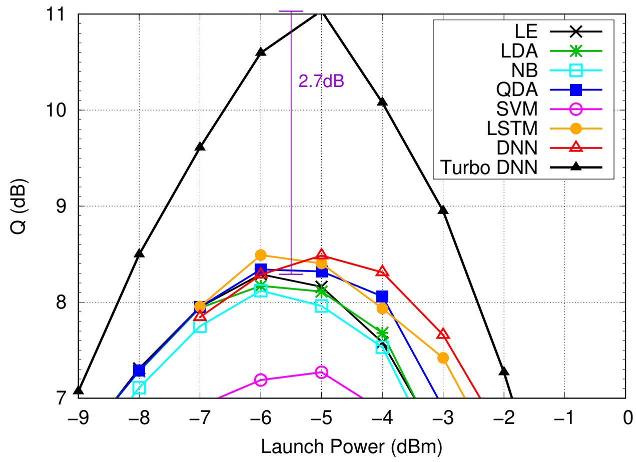


Fig. 6: Single-iteration DNN-TEQ for DP-4QAM 50-span NZDSF (DVB-S2 LDPC code of rate $9/10$).

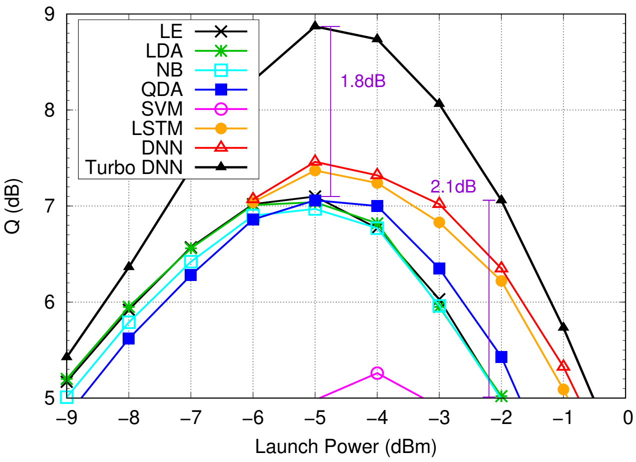


Fig. 7: Single-iteration DNN-TEQ for DP-16QAM 16-span NZDSF (DVB-S2 LDPC code of rate $8/9$).

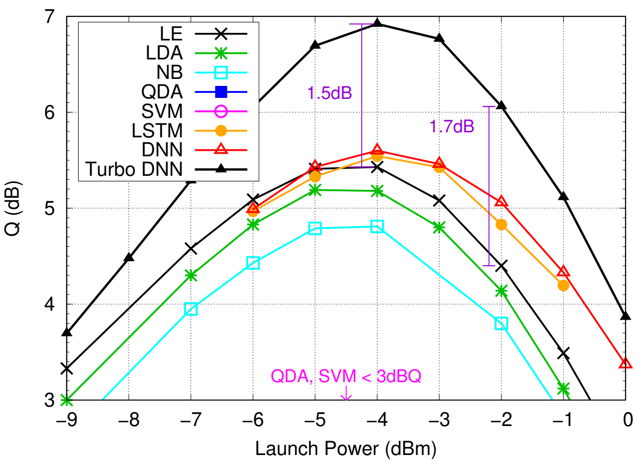


Fig. 8: Single-iteration DNN-TEQ for DP-64QAM 8-span NZDSF (DVB-S2 LDPC code of rate $4/5$).

6 References

[1] I. B. Djordjevic, L. L. Minkov, and H. G. Batshon, "Mitigation of linear and nonlinear impairments in high-speed optical networks by using LDPC-coded turbo equalization," *JSAC* **26** 6 (2008), pp. 73–83.

[2] H. G. Batshon, I. B. Djordjevic, L. Xu, and T. Wang, "Iterative polar quantization based modulation to achieve channel capacity in ultra-high-speed optical communication systems," *IEEE Photon. Journal* **2** 4 (2010), pp. 593–9.

[3] C. Duan, K. Parsons, T. Koike-Akino, R. Annavajjala, K. Kojima, T. Yoshida, T. Sugihara, and T. Mizuochi, "A low-complexity sliding-window turbo equalizer for nonlinearity compensation," *OFC* (2012), JW2A.59.

[4] F. P. Guiomar, S. B. Amado, N. J. Muga, J. D. Reis, A. L. Teixeira, and A. N. Pinto, "Simplified Volterra series nonlinear equalizer by intra-channel cross-phase modulation oriented pruning," *ECOC* (2013), We3C6.

[5] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *JLT* **26** 20 (2008), pp. 3416–25.

[6] W. Yan, Z. Tao, L. Dou, L. Li, S. Oda, T. Tanimura, T. Hoshida, and J. C. Rasmussen, "Low complexity digital perturbation back-propagation," *ECOC* (2011), Tu.3.A.2.

[7] N. Irukulapati, H. Wymeersch, and P. Johannisson, "Extending digital backpropagation to account for noise," *ECOC* (2013), We.3.C.4.

[8] C. Häer and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," *OFC* (2018), W3A.4.

[9] T. Koike-Akino, "Perspective of statistical learning for nonlinear equalization in coherent optical communications," *SPPCom* (2014), ST2D-2.

[10] T. Koike-Akino, C. Duan, K. Parsons, K. Kojima, T. Yoshida, T. Sugihara, and T. Mizuochi, "High-order statistical equalizer for nonlinearity compensation in dispersion-managed coherent optical communications," *Opt. Expr.* **20** 14 (2012), pp. 15769–80.

[11] X. Xie, F. Yaman, X. Zhou, and G. Li, "Polarization demultiplexing by independent component analysis," *IEEE PTL* **22** 11 (2010), pp. 805–7.

[12] T. Koike-Akino, K. Kojima, D. S. Millar, K. Parsons, Y. Miyata, W. Matsumoto, and T. Mizuochi, "Cycle slip-mitigating turbo demodulation in LDPC-coded coherent optical communications," *OFC* (2014), M3A-3.

[13] E. Giacoumidis, S. Mhatli, T. Nguyen, S. T. Le, I. Aldaya, M. McCarthy, and B. Eggleton, "Kerr-induced nonlinearity reduction in coherent optical OFDM by low complexity support vector machine regression-based equalization," *OFC* (2016), Th2A-49.

[14] M. A. Jarajreh, E. Giacoumidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE PTL* **27** 4 (2015), pp. 387–90.

[15] E. Giacoumidis, S. T. Le, M. Ghanbarisabagh, M. McCarthy, I. Aldaya, S. Mhatli, M. A. Jarajreh, P. A. Haigh, N. J. Doran, A. D. Ellis, and B. J. Eggleton, "Fiber nonlinearity-induced penalty reduction in CO-OFDM by ANN-based nonlinear equalization," *Opt. Lett.* **40** (2015), pp. 5113–6.

[16] R. Rios-Müller, J. M. Estarán, and J. Renaudier, "Experimental estimation of optical nonlinear memory channel conditional distribution using deep neural networks," *OFC* (2017), W2A5.1.

[17] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, and Y. Inada. "Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable," *OFC* (2018), Th4D-5.

[18] P. Li, L. Yi, L. Xue, and W. Hu, "56 Gbps IM/DD PON based on 10G-class optical devices with 29 dB loss budget enabled by machine learning," *OFC* (2018), M2B.2.

[19] T. Koike-Akino, D. S. Millar, K. Parsons, and K. Kojima, "Fiber nonlinearity equalization with multi-label deep learning scalable to high-order DP-QAM," *SPPCom* (2018), SpM4G.1.

[20] C. Y. Chuang, C. C. Wei, T. C. Lin, K. L. Chi, L. C. Liu, J. W. Shi, Y. K. Chen, and J. Chen, "Employing deep neural network for high speed 4-PAM optical interconnect," *ECOC* (2017).

[21] C. Ye, D. Zhang, X. Hu, X. Huang, H. Feng, and K. Zhang, "Recurrent neural network (RNN) based end-to-end nonlinear management for symmetrical 50Gbps NRZ PON with 29dB+ loss budget," *ECOC* (2018).

[22] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. B ulow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *JLT* **36** 20 (2018), pp. 4843–55.

[23] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," *ECOC* (2018),

[24] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities," *ECOC* (2018).

[25] M. Chagnon, B. Karanov, and L. Schmalen, "Experimental demonstration of a dispersion tolerant end-to-end deep learning-based IM-DD transmission system," *ECOC* (2018).

[26] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE TCOM* **52** 4 (2004), pp. 670–8.