SubOptic2019
TO THE BEACH AND BEYOND
RETHINKING GLOBAL NETWORKS

FIBER NONLINEARITY COMPENSATION BY NEURAL NETWORKS

Shaoliang Zhang[(1)], Fatih Yaman[(1)], Takanori Inoue[(2)], Kohei Nakamura[(2)], Eduardo Mateo[(2)], Yoshihisa Inada[(2)]

[(1)] NEC Laboratories America, Inc., Princeton, NJ 08540, USA, szhang@nec-labs.com
[(2)] Submarine Network Division, NEC Corporation, Tokyo 108-8001, Japan

Abstract: Neuron network (NN) is proposed to work together with perturbation-based nonlinearity compensation (NLC) algorithm by feeding with intra-channel cross-phase modulation (IXPM) and intra-channel four-wave mixing (IFWM) triplets. Without prior knowledge of the transmission link and signal pulse shaping/baudrate, the optimum NN architecture and its tensor weights are completely constructed from a data-driven approach by exploring the training datasets. After trimming down the unnecessary input tensors based on their weights, its complexity is further reduced by applying the trained NN model at the transmitter side thanks to the limited alphabet size of the modulation formats. The performance advantage of Tx-side NN-NLC is experimentally demonstrated using both single-channel and WDM-channel 32Gbaud dual-polarization 16QAM over 2800km transmission.

1. INTRODUCTION

Digital coherent transmission systems nowadays play a key role to support capacity-demanding backbone networks to meet the increasing internet traffic [1]. Although advanced modulation formats are capable of increasing the spectral usage given the limited optical bandwidth in each fiber [2, 3], the maximum achievable information rate of long-haul transmission systems is mainly limited by the fiber's Kerr nonlinear interaction between signal intensity and phase [4]. Nonlinearity compensation (NLC) algorithm has been proposed to mitigate the fiber nonlinearity which can be characterized by nonlinear Schrödinger equation (NLSE). Its solutions can be approximated by single-step and multiple-step methods. Many NLC algorithm has been proposed, such as perturbation-based pre/post-distortion (PPD) [5], Volterra equalizers [6] and digital backpropagation (DBP) [7]. However, all the proposed NLC algorithm requires the accurate knowledge of the transmission link parameters, such as dispersion, fiber nonlinearity and span length, which may be not available in a software-defined meshed network. Recently, an adaptive approach based on filtered-DBP algorithm has been proposed to address this issue [8]. Alternatively, a machine-learning algorithm was used to predict the received signal nonlinearity without prior knowledge of the link parameters from the received symbol triplets [9]. The intra-channel cross-phase modulation (IXPM) and intra-channel four-wave mixing (IFWM) triplets are fed into a neuron network (NN) to explore the correlation among these triplets. In this paper, the derived NN-NLC model is applied at the transmitter side to pre-distort the signal symbols. The complexity of the proposed NN-NLC algorithm is experimentally shown to have performance advantage over filtered-DBP [10] at the same computation complexity in terms of real multiplications per symbol.

2. THE ARCHITECURE OF NN-NLC

The proposed NN-NLC algorithm consists of training and execution stages. During the training stage, it operates on the soft data from carrier phase recovery in the receiver's DSP. The performance of the model is checked against the cross-validation (CV) dataset only to optimize the NN model and its tensors' parameters, such as weights and biases. Afterwards the learned model is

applied to the uncorrelated test dataset at either transmitter or receiver side for all channel powers in the execution stage.

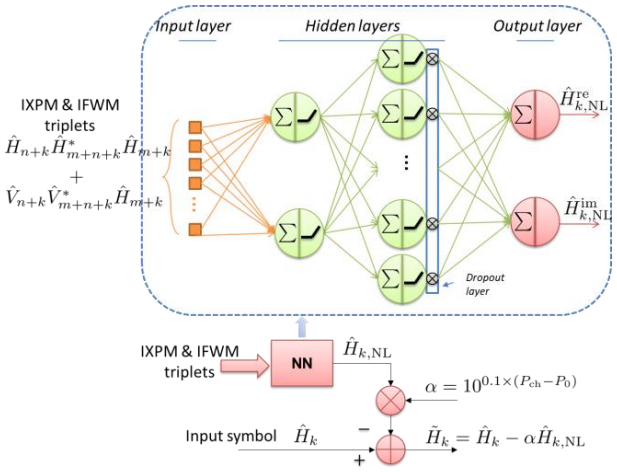

Fig. 1 The block diagram of the proposed NN-NLC illustrated for pol-H only. The diagram in the dashed box describes the optimized NN architecture with 2 hidden layers used in the paper.

A fully-connected neuron network shown in Fig. 1 is constructed in TensorFlow with an input layer with $2N_t$ triplets nodes, 2 hidden layers consisting of 2 and 10 nodes, respectively, and two output nodes corresponding to the real and imaginary of the estimated nonlinearity. Note that these triplets are separated into real and imaginary before being fed into the NN model. The activation function Leaky-ReLu is applied in the nodes of both hidden layers [11]. A dropout layer with probability of 0.5 is placed after the 2nd hidden layer during training only to avoid overfitting. The built-in *Adam* learning algorithm with a learning rate of 0.001 and batch size of B = 100 are used to train the NN by transmitting known but randomly generated patterns, and searching for the best node tensor parameters that minimize the mean square error (MSE) between the transmitted and received symbols after NN-LC, given by

$$\mathrm{MSE} = \frac{1}{B}\sum_{i=1}^{B}\left|H_i - (\hat{H}_i - \hat{H}_{i,\mathrm{NL}})\right|^2,$$

where $\hat{H}_i$ and $\hat{H}_{i,NL}$, respectively, are the received symbols and estimated nonlinearity for pol-H, H_i is the transmitted symbol, and $|\cdot|$ stands for absolute operation. The same notation is also used for V_i in V polarization. Although the model is trained using pol-H data, the similar performance improvement is observed for the pol-V when applying the same model. Note that the training can be done at much slower pace than data rate to allow deep-learning algorithm to locate the appropriate NN models and compute the optimum tensor weights prior to the execution stage.

In order to more accurately characterize the fiber nonlinearity, signal nonlinearity has to be observed in the received training data. The launch channel power P_0 can be set beyond the optimum channel power to allow nonlinearity noise dominant over received amplified spontaneous emission (ASE) noises. In addition, de-noising averaging, such as averaging over multiple datasets, can be carried out for the fixed training data pattern to isolate the data-dependent nonlinearity from additive Gaussian ASE noises. After cleaning up the ASE noises in the received training dataset, these data are ready to be used for computing the IXPM & IFWM triplets $\hat{H}_{n+k}\hat{H}^*_{m+n+k}\hat{H}_{m+k}$ and $\hat{V}_{n+k}\hat{V}^*_{m+n+k}\hat{H}_{m+k}$. Here, the subscripts m and n are selected from $-\lfloor L/2\rfloor \le m,n \le \lceil L/2\rceil$-1, and stand for the symbol index with respect to the center symbols H_k and V_k in a symbol window length of L. Here $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote rounding downwards and upper towards the nearest integer.

In previous work, the nonlinear perturbation coefficients $C_{m,n}$ associated with each triplets are first analytically computed based on the link parameters and signal baud rate, and only those triplets with coefficients above a certain threshold κ, i.e., $20\log_{10}\left|\frac{C_{m,n}}{C_{0,0}}\right| > \kappa$, are retained for feeding into the NN models [9]. Due to the hyperbola characteristic of the nonlinear perturbation coefficients $C_{m,n}$ at the

given *m*, here we propose to select only those triplets based on the criteria:

$$|n| \le \min\left\{\frac{\rho\lceil L/2\rceil}{|m|}, \left\lceil\frac{L}{2}\right\rceil\right\},$$

where ρ is a control factor to select the number of triplets N_t, and $\min\{\cdot\}$ takes the minimum of its arguments. As a result, the prior information about the link parameters is completely not necessary for NN-NLC algorithm.

Given the symbol of interest H_k centered at the middle of symbol length L, the IXPM and IFWM triplets are passed to the NN model in Fig.1 to estimate the nonlinearity. The estimated $\widehat{H}_{k,NL}$is first scaled by the channel power P_{ch} of the test dataset with respect to the reference channel power P_0 of the training data used for deriving the model, i.e., $10^{0.1\times(P_{ch}-P_0)}$, and are then subtracted from the original received symbol $\widehat{H}_k$ to mitigate the nonlinearity.

3. EXPERIMENTS

The system setup is shown in Fig. 2a. The testing 32 Gbaud dual-polarization (DP) 16 quadrature amplitude modulation (16QAM) signal with root-raised cosine (RRC) 0.01 pulse shaping is generated by a DP-IQ modulator driven by 64Gs/s digital-to-analog converter (DAC). The wavelength-division-multiplexing (WDM) dummies are separated into odd and even groups, and each is generated using single-polarization IQ modulator and polarization multiplexer (PolMux) to have independent PolMux-16QAM dummy channels at RRC 0.01 shaping. The combined WDM signals are inserted into the gap of ASE dummy opened up by the wavelength-selective switch (WSS), and are then sent to the loop test bed, which is comprised of five spans of 80 km single-mode fiber (SMF) with 0.2 dB/km loss and 17 ps/nm/km chromatic dispersion (CD). Gain equalizer (GEQ) and polarization scrambler (PS) are used to balance the EDFA gain tilt and to avoid spurious polarization effects from recirculation. A digital coherent

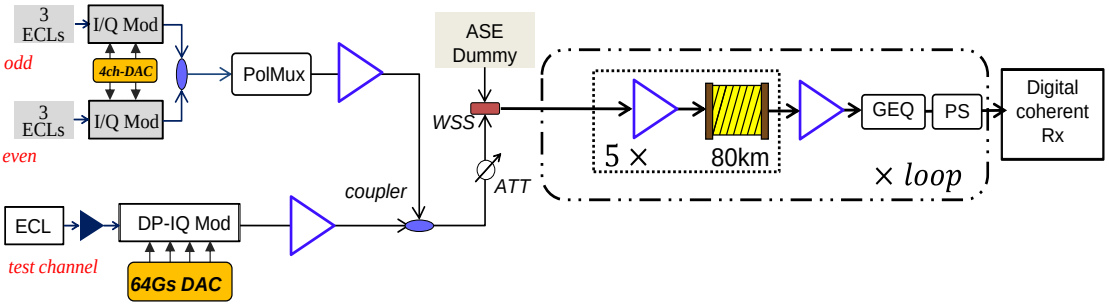

Fig.2 (a) System setup of the transmitter and transmission loop. DAC: digital-to-analog converter, ECL: external cavity laser, GEQ: gain equalizer, PolMux: polarization multiplexer, PS: polarization scrambler, WSS: wavelength-selective switch.

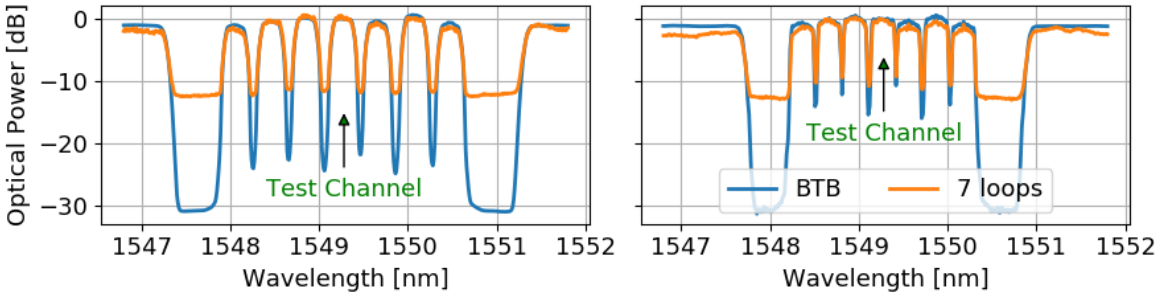

Fig.2 (b) WDM Spectra at 50GHz (left) and 37.5GHz (right)

receiver running at 50-GSa/s with analog bandwidth of 20 GHz down samples the optical waveforms for offline DSP algorithm to recover the transmitted symbols, including chromatic dispersion compensation (CDC), decision-directed (DD) least-mean-square (LMS) polarization demuliplexer and carrier phase recovery. In addition, 50% CDC has been applied at the transmitter to reduce the interaction lengths [12]. Three uncorrelated datasets with ~115k symbols each are generated for training, CV and testing. To ensure data independence, the data pattern used in the training, CV and test datasets is measured to have maximum 0.6% normalized cross-correlation. Fig.2b plot the WDM spectra at transmitter and receiver after 7 loops.

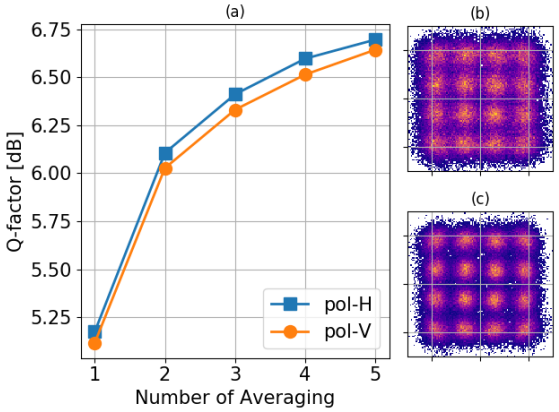

Fig. 3 The impact of de-noise averaging of the training datasets on the Q-factor at SNR = 18.4 dB after 2800 km

Single-channel 32Gbaud DP-16QAM experimental data is first used to elaborate the proposed NN-NLC algorithm. Multiple waveform acquisitions are processed, and the recovered soft symbols after carrier phase recovery are aligned to average out the additive noise. Fig. 3 plots the impact of the number of acquired waveforms on the Q-factor and constellation of the training dataset received at ~2 dB higher channel power than the optimum after 2800 km transmission. About 1.6 dB Q-factor improvement has been observed after only averaging over 5 acquired waveforms. The saturation curves show that the resulting cleaner constellation in Fig.3c is able to more accurately characterize the nonlinear noise than the one in Fig.3b. Using the de-noised training data with ρ = 1, L = 151 and N_t = 1929, the density map of the tensor weights $W_{m,n}$ at the input layer after training is plotted in Fig. 4(a). As can be seen, the triplets selection criterion in Eq(2) allows the deep-learning algorithm to efficiently locate the important triplets rather than exploring all L^2 triplets, thus significantly reducing the computation complexity and training time.

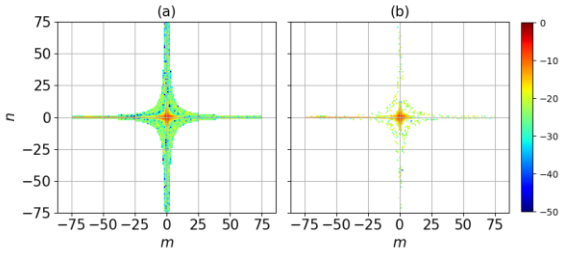

Fig. 4 The density plot of the input layer weights of the NN model at initial N_t = 1929 (a) and N_t = 615 triplets after iterative trimming (κ = -22dB) (b).

As shown in Fig.4a, with the initial N_t = 1929 triplets, some of the input tensor weights $W_{m,n}$ in the trained model show much smaller contribution to the signal nonlinearity than the center ones. As a result, the number of triplets N_t can be further reduced by only keeping those weights larger than a threshold κ, i.e., $20\log_{10}\left|\frac{W_{m,n}}{W_{0,0}}\right| > \kappa$. After trimming off the weights $W_{m,n}$ that are less than κ = -22dB shown in Fig.4a, the remaining 615 triplets are re-trained in the NN model and the new density plot of the input tensor weights $W_{m,n}$ is shown in Fig.4b. Fig. 5 plots the impact of the trimming threshold κ on the performance improvement of the NN-NLC as a function of received SNR after 2800km transmission. At the optimum received SNR, the NN-NLC algorithm at trimming threshold κ < -15 dB achieves > 0.5 dB Q improvement over CDC. The more Q improvement at the highest received SNR further confirms the NN model is able to accurately predict the signal nonlinearity. By adjusting the trimming threshold κ from -

35dB to -15dB, zoom inset of Fig. 5 shows the performance trade-off within ~0.2dB and ~0.4dB Q variation, respectively, at received SNR of 16.6 dB and 18.4 dB.

The transmitter (Tx)-side NN-NLC has the advantage to avoid the computation of triplets using LUT to reduce the complexity. In addition, since the NN model works on the clean transmitted symbol, the Q improvement of NN-NLC over CDC could be higher than the one at the receiver side. The third advantage is that the receiver DSP algorithm works on the signals with less nonlinearity, thus likely to reduce cycle slip rate [13]. Applying the NN model derived with trimming threshold κ = -22dB at the receiver side to the original 16QAM symbols in the test dataset, the pre-distorted symbol constellation is plotted in Fig. 6a. Compared to the recovered constellation without NN-NLC algorithm shown in Fig.6c, the transmitter-side NN-NLC can significantly improve the constellation quality as plotted in Fig.6b. Compared to the Rx-side NN-NLC at the trimming threshold κ = -22dB, Fig. 6 shows ~0.1 dB Q improvement when moving the NN-NLC to the transmitter side thanks to the undistorted transmitter symbols.

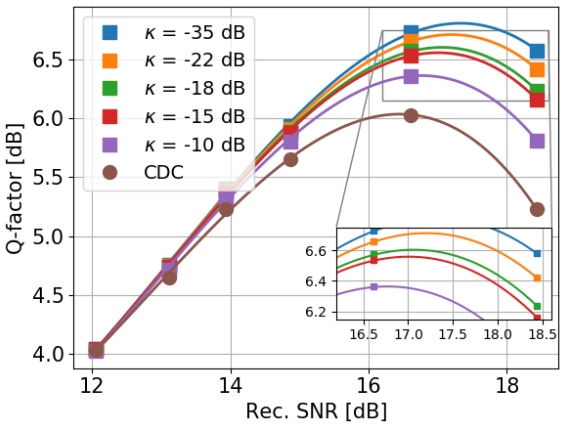

Fig. 5 The impact of trimming threshold κ on the NN-NLC at the receiver side after 2800km transmission.

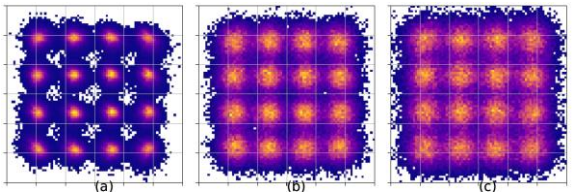

Fig. 6 Pre-distorted symbols at transmitter side (a); recovered constellation at the receiver with (b) and without (c) transmitter-side NN-NLC. Received SNR = 18.4 dB@2800 km transmission.

Since the complexity of real multiplications could be 4 times as much as addition operation [14], only real multiplication will be taken into account when comparing the complexity of the NLC algorithm. The NN model shown in Fig.1 requires $4N_t + 40$ real multiplications because of three cross-layer tensor interaction. Note that the activation function Leaky-ReLU in the hidden nodes and IXPM/IFWM triplets computation are assumed to be implemented in LUT. After scaling the estimated nonlinearity term, the number of real multiplication per symbol for the proposed NN-ALC shown in Fig.1 is summarized as $4N_t + 42$. Therefore, reducing the number of triplets N_t is proven to be the most effective way to lower the complexity of the NN-NLC algorithm in our model.

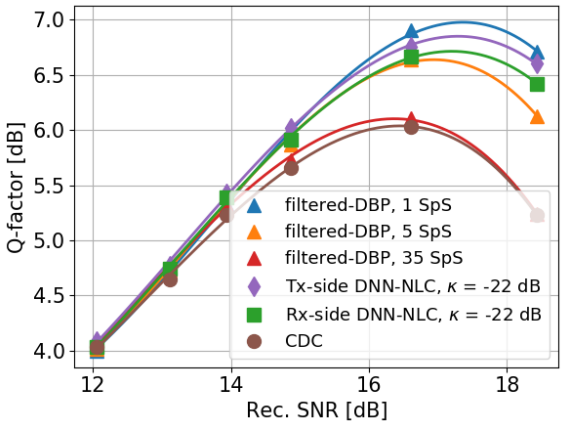

Fig. 7 The performance comparison between NN-NLC at transmitter and receiver side, and filtered-DBP at different spans per step for single-channel DP-16QAM after 2800 km

On the other hand, filtered-DBP is a well-known technique in the literature to balance the trade-off between performance and computation complexity. Since multiple spans are emulated in each DBP step, the intensity waveforms has to be filtered by a Gaussian low-pass filter (LPF) prior to de-rotating signal phase [15]. Symmetric split-step method is used in our experiments, and the optimal bandwidth is found to be ~5 GHz, ~1 GHz, ~1 GHz. ~0.8 GHz, ~0.8 GHz and ~0.5 GHz for 1, 5, 7, 12, 18, and 35 spans per step (SpS). The optimum scaling factor ξ used to de-rotate the signal's phase is about 0.7 for all cases. Compared to single-step NN-NLC, Fig. 7 shows that single-step filtered-DBP (35 SpS) is outperformed by > 0.6 dB. From the measurement results, filtered-DBP needs at least 7 steps, i.e., 5 SpS, to catch up the performance of Rx-side NN-NLC at κ= -22dB.

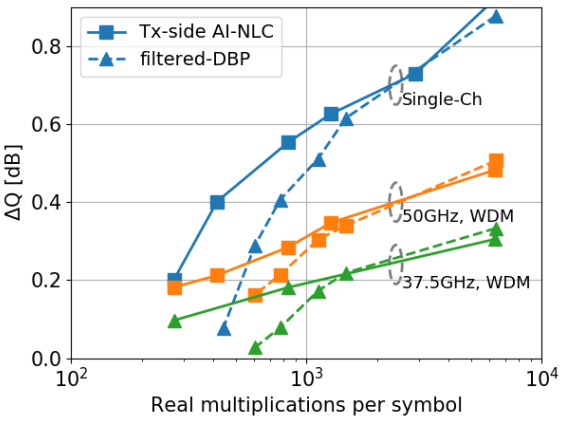

Fig. 8 The comparison of Q-factor improvement between Tx-side NN-NLC and filtered-DBP. FFT size is assumed to be 4086.

The Q performance improvement over CDC is plotted in Fig.8 versus real multiplications per symbol for filtered-DBP and Tx-side NN-NLC algorithm at both single- and WDM-channel cases. In general, Tx-side NN-NLC performs better than filtered-DBP only when the computation complexity is less than ~2000 real multiplications per symbol. The performance of Tx-side NN-NLC is able to match the performance of the filtered-DBP even at higher complexity up to 1 SpS. The Q-factor improvement both NN-NLC and filtered-DBP over CDC drops to 0.3 dB, 0.5 dB, respectively, at 37.5 GHz and 50 GHz WDM from ~0.9 dB at single channel. The mitigation of inter-channel nonlinearity needs to be addressed to see further improvement at affordable complexity.

4. CONCLUSION

In this paper, a data-driven neural network (NN) is demonstrated to predicting the fiber nolinearity from the IXPM/IFWM triplets. Without prior knowledge of the fiber link, the estimated fiber nonlinearity is further used in the system-agnostic NN-NLC algorithm to mitigate the signal nonlinearity and is experimentally demonstrated to have ~0.9 dB, ~0.5 dB and ~0.3 dB Q-factor improvement over CDC. The complexity of NN-NLC algorithm is further reduced by trimming the input tensor weights and being applied at the transmitter side. Based on the 32Gbaud DP-16QAM at both single-channel and WDM-channel over 2800km transmission, Tx-side NN-NLC is able to outperform the filtered-DBP algorithm when the computation complexity is less than ~2000 real multiplications per symbol.

5. REFERENCES

[1] K. Roberts, Q. Zhuge, I. Monga, S. Gareau, and C. Laperle, "Beyond 100Gb/s: Capacity, flexibility, and network optimization," J. Opt. Commun. Netw. 9, C12–C24 (2017).

[2] A. Ghazisaeidi, I. F. de Jauregui Ruiz, R. Rios-MÃ¼ller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet, and J. Renaudier, "Advanced C+L-band transoceanic transmission systems based on probabilistically shaped pdm-64qam," J. Light. Technol. 35, 1291–1299 (2017).

[3] S. Zhang, F. Yaman, Y.-K. Huang, J. D. Downie, X. Sun, A. Zakharian, R. Khrapko, W. Wood, I. B. Djordjevic, E. Mateo, and Y.

Inada, "50.962Tb/s over 11185 km Bi-directional C+L transmission using optimized 32QAM," in Conference on Lasers and Electro-Optics, (Optical Society of America, 2017), p. JTh5A.9.
[4] R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," J. Light. Technol. 28, 662–701 (2010).
[5] Z. Tao, L. Dou, W. Yan, L. Li, T. Hoshida, and J. C. Rasmussen, "Multiplier-free intrachannel nonlinearity compensating algorithm operating at symbol rate," J. Light. Technol. 29, 2570–2576 (2011).
[6] F. P. Guiomar, J. D. Reis, A. L. Teixeira, and A. N. Pinto, "Mitigation of intra-channel nonlinearities using a frequency-domain volterra series equalizer," Opt. Express 20, 1360–1369 (2012).
[7] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," J.Light. Technol. 26, 3416–3425 (2008).
[8]F. Zhang, Q. Zhuge, M. Qiu, X. Zhou, M. Y. S. Sowailem, T. M. Hoang, M. Xiang, and D. V. Plant, "Blind adaptive digital backpropagation for fiber nonlinearity compensation," J. Light. Technol. 36, 1746–1756 (2018).
[9] 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," in Optical Fiber Communication Conference Postdeadline Papers, (Optical Society of America, 2018), p. Th4D.5.
[10] L. B. Du and A. J. Lowery, "Improved single channel backpropagation for intra-channel fiber nonlinearity compensation in long-haul optical communication systems," Opt. Express 18, 17075–17088 (2010).
[11] Goodfellow, I., Bengio, Y. & Courville, A. Deep Learning (MIT Press, 2016).
[12] Y. Gao, J. C. Cartledge, A. S. Karar, S. S.-H. Yam, M. O'Sullivan, C. Laperle, A. Borowiec, and K. Roberts, "Reducing the complexity of perturbation based nonlinearity pre-compensation using symmetric EDC and pulse shaping," Opt. Express 22, 1209–1219 (2014).
[13] H. Zhang, Y. Cai, D. G. Foursa, and A. N. Pilipetskii, "Cycle slip mitigation in polmux-qpsk modulation," in 2011 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference, (2011), pp. 1–3.
[14] T. Oyama, T. Hoshida, H. Nakashima, T. Tanimura, Z. Tao, and J. C. Rasmussen, "Proposal of improved 16qam symbol degeneration method for simplified perturbation-based nonlinear equalizer," in 2014 OptoElectronics and Communication Conference and Australian Conference on Optical Fibre Technology, (2014), pp. 941–943.
[15] Y. Gao, J. H. Ke, K. P. Zhong, J. C. Cartledge, and S. S. H. Yam, "Assessment of intrachannel nonlinear compensation for 112 gb/s dual-polarization 16qam systems," J. Light. Technol. 30, 3902–3910 (2012).

