Neural Turbo Equalization: Deep Learning for Fiber-Optic Nonlinearity Compensation

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

TR2020-033 April 03, 2020

Abstract

Recently, data-driven approaches motivated by modern deep learning have been applied to optical communications in place of traditional model-based counterparts. The application of deep neural networks (DNN) allows flexible statistical analysis of complicated fiber-optic systems without relying on any specific physical models. Due to the inherent nonlinearity in DNN, various equalizers based on DNN have shown significant potentials to mitigate fiber nonlinearity. In this paper, we propose turbo equalization (TEQ) based on DNN as a new alternative framework to deal with nonlinear fiber impairments. The proposed DNN-TEQ is constructed with nested deep residual networks (ResNet) to train extrinsic likelihood given soft-information feedback from channel decoding. Through extrinsic information transfer (EXIT) analysis, we verify that our DNN-TEQ can accelerate decoding convergence to achieve a significant gain in achievable throughput by 0.61 b/s/Hz. We also demonstrate that optimizing irregular low-density parity-check (LDPC) codes based on the EXIT chart of the DNN-TEQ can improve achievable rates by up to 0.12 b/s/Hz.

Journal of Lightwave Technology

Neural Turbo Equalization: Deep Learning for Fiber-Optic Nonlinearity Compensation

Toshiaki Koike-Akino, *Senior Member, IEEE, Senior Member, OSA*, Ye Wang, *Senior Member, IEEE*, David S. Millar, *Member, IEEE, Member, OSA*, Keisuke Kojima, *Senior Member, IEEE, Fellow, OSA*, Kieran Parsons, *Senior Member, IEEE, Member, OSA*,

***Abstract*—Recently, data-driven approaches motivated by modern deep learning have been applied to optical communications in place of traditional model-based counterparts. The application of deep neural networks (DNN) allows flexible statistical analysis of complicated fiber-optic systems without relying on any specific physical models. Due to the inherent nonlinearity in DNN, various equalizers based on DNN have shown significant potentials to mitigate fiber nonlinearity. In this paper, we propose turbo equalization (TEQ) based on DNN as a new alternative framework to deal with nonlinear fiber impairments. The proposed DNN-TEQ is constructed with nested deep residual networks (ResNet) to train extrinsic likelihood given soft-information feedback from channel decoding. Through extrinsic information transfer (EXIT) analysis, we verify that our DNN-TEQ can accelerate decoding convergence to achieve a significant gain in achievable throughput by 0.61 b/s/Hz. We also demonstrate that optimizing irregular low-density parity-check (LDPC) codes based on the EXIT chart of the DNN-TEQ can improve achievable rates by up to 0.12 b/s/Hz.**

***Index Terms*—Deep Learning, turbo equalization, digital signal processing, fiber nonlinearity, high-order QAM, LDPC codes**

I. INTRODUCTION

MACHINE learning techniques [1]–[3] have been recently applied to optical communications systems to deal with various issues such as network monitoring [4]–[6], traffic control [7]–[10], signal design [11]–[15], and nonlinearity compensation [16]–[21]. Since the fiber nonlinearity is a major limiting factor to the achievable information rates [22]–[24], mitigating nonlinearity has been of great importance to realize reliable, high-speed, and long-reach optical communications. Conventionally, a number of model-based nonlinear equalizers to compensate for fiber distortion were investigated, e.g., maximum-likelihood sequence equalizer (MLSE) [25]–[27], turbo equalizer (TEQ) [28]–[30], Volterra series transfer function [32], [33], and digital backpropagation (DBP) [35]–[38]. However, those nonlinear equalizations are computationally complex and susceptible to model parameter mismatch in general. Recent data-driven approaches motivated by deep learning can favorably replace such traditional model-based methods as the use of deep neural networks (DNN) allows flexible statistical analysis of complicated fiber-optic systems without relying on specific models. In the past few years, DNN has shown its high potential in nonlinear performance improvement, e.g., [12]–[21].

Nonetheless, most existing 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 is relevant only when nonbinary FEC codes are assumed. For more practical bit-interleaved coded modulation (BICM) systems, it was found in [20] that binary cross-entropy (BCE) loss can improve accuracy and scalability to high-order quadrature-amplitude modulation (QAM). In this paper, we propose a novel 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 framework of BICM-ID which takes soft-decision messages from the FEC decoder to refine the DNN output for improved equalization accuracy. We analyze the extrinsic information transfer (EXIT) of turbo DNN, and demonstrate that the proposed DNN paired with irregular low-density parity-check (LDPC) codes used in DVB-S2 standards offers a significant performance gain by accelerating the decoder convergence in nonlinear transmissions.

The contributions of this paper are summarized as follows:

- **Trend overview**: We overview the recent trend of deep learning in optical communications literature.
- **Multi-label DNN**: We verify that nonbinary cross-entropy is not scalable to high-order QAM signals and DNN trained with BCE loss can appropriately compensate for fiber nonlinearity.
- **Turbo DNN**: We propose a nested residual DNN architecture for TEQ to further improve performance.
- **EXIT analysis**: We analyze the EXIT chart of our DNN-TEQ and show that DNN-TEQ accelerates decoding convergence.
- **LDPC design**: We optimize the degree distribution of LDPC codes based on the EXIT charts of DNN-TEQ, achieving higher throughput.

Note that due to the above contributions, in particular the demonstration of rate improvement with optimized LDPC codes for DNN-TEQ, this paper is distinguished from our preliminary reports [20], [21], [48]. To the best of our knowledge,

Manuscript received November 15, 2019; revised January 29, 2020; accepted xxxx.

The authors are with Mitsubishi Electric Research Laboratories (MERL), 201 Broadway, Cambridge, MA 02139, USA (e-mail: koike@merl.com; yewang@merl.com; millar@merl.com; kojima@merl.com; parsons@merl.com).

This paper contains in part our previous work [20], [21], [48].

Color versions of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier xx.xxx/JLT.2020.xxx

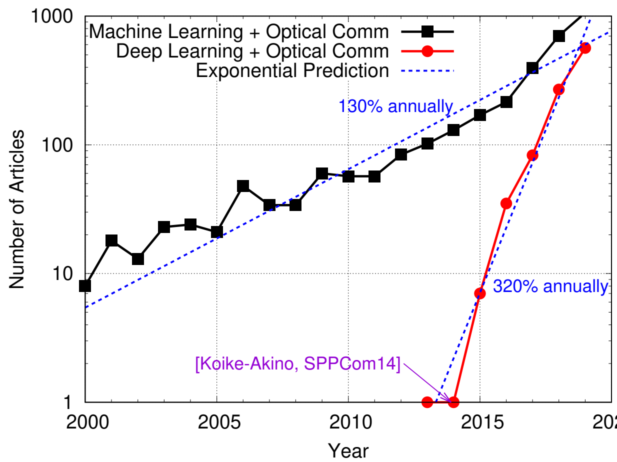


Fig. 1. Machine/deep learning trend in optical communication applications (keyword hits on Google Scholar, excluding non-relevant ones).

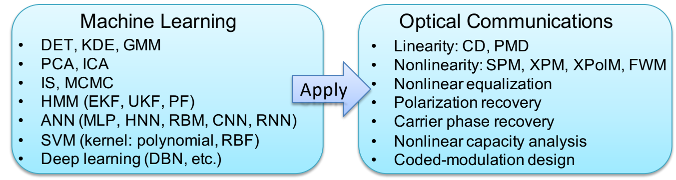


Fig. 2. Machine learning approaches applied to optical communications [48].

there is no other literature which applied DNN to TEQ for nonlinear compensation.

II. Machine Learning for Optical Communications

A. Trend Overview

Fiber-optic communications suffer from various linear and nonlinear impairments, such as amplified spontaneous emission (ASE) noise, laser linewidth, chromatic dispersion (CD), polarization mode dispersion (PMD), self-phase modulation (SPM), cross-phase modulation (XPM), cross-polarization modulation (XPolM), and four-wave mixing (FWM) [22]–[24]. Although the physics is well captured by the nonlinear Schrödinger equation model, the high-complexity split-step Fourier method is required for solving lightwave propagation numerically. It is hence natural to admit that the nonlinear physics calls for nonlinear signal processing to appropriately deal with the nonlinear distortions in practice.

In place of conventional model-based nonlinear signal processing, the application of machine learning techniques [1]–[3] to optical communication systems has recently received increased attention [4]–[21]. The promise of such data-driven approaches is that learning a black-box DNN could potentially overcome situations where limited models are inaccurate and complex theory is computationally intractable.

Fig. 1 shows the trend of machine learning applications in optical communications society in the past two decades. Here, we plot the number of articles in each year according to Google Scholar search of the keyword combinations; "machine learning" + "optical communication" or "deep learning" + "optical communication." As we can see, machine learning has been already used for optical communications since twenty years ago. Interestingly, we observe an exponential trend in which the number of applications exponentially grows by a factor of nearly 130% per year. For deep learning applications, more rapid annual increase by a factor of 320% can be found in the past half decade. As of today, there are nearly a thousand articles on deep learning applications. Note that the author's article [48] in 2014 is one of very first papers discussing the application of deep learning to optical communications.

B. Machine Learning Techniques

We briefly overview some learning techniques to analyze nonlinear statistics applied to optical communications as shown in Fig. 2. For example, kernel density estimation (KDE), density estimation trees (DET), and Gaussian mixture model (GMM) are alternatives to histogram analysis. Principal component analysis (PCA) and independent component analysis (ICA) are used to analyze major factors of data. For high-dimensional data sets, we may use Markov-chain Monte–Carlo (MCMC) and importance sampling (IS). To analyze stochastic sequence data, extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filter (PF) based on hidden Markov model (HMM) may be used.

Since the mid-70's, artificial neural networks (ANN) have been a major theme in machine learning research. Various architectures including restricted Boltzmann machines (RBM), multi-layer perceptron (MLP), Hopfield neural networks (HNN), convolutional neural networks (CNN), and recurrent neural networks (RNN) have been investigated. Since the mid-90's, the support vector machine (SVM) emerged as an influential method to address nonlinear statistics via the kernel trick, which analyzes higher-dimensional linearlized feature spaces called reproducing kernel Hilbert space with kernel functions, such as the radial basis function (RBF). Since 2006, deep learning [1] based on DNN has been a major breakthrough in media signal processing fields. In deep learning, many-layer deep belief networks (DBN) are trained with massively large amount of data.

C. Classical Machine Learning Applications

We show a few examples of machine learning approaches applied to nonlinear fiber-optic communications. In [39], the use of ICA is proposed for polarization recovery as an alternative to constant-modulus adaptation. Shallow ANN-based nonlinear equalizers have been studied in the literature [40]–[42]. We have investigated GMM-based TEQ and MLSE receivers [27], where 2 dB performance gain was achieved over DBP. SVM has been used as another equalizer [43], [44], in which an irregular decision rule like Yin–Yang spiral boundary [45] can be learned by kernel-SVM. RBF kernels have been studied in other literature, e.g., [46]. HMM-based cycle-slip compensation [47] offers greater than 2 dB gain. A stochastic DBP proposed in [38] exhibits an outstanding performance by adopting MCMC particle representation of stochastic noise.

D. Modern Deep Learning Applications

As shown in Fig. 1, there exist a lot of deep learning applications, among which a limited number of examples are

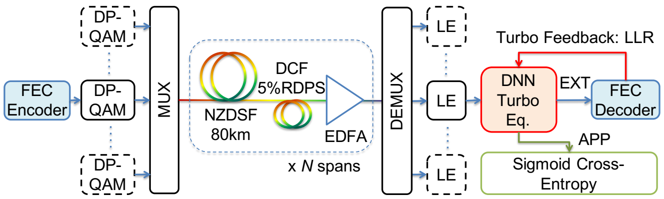
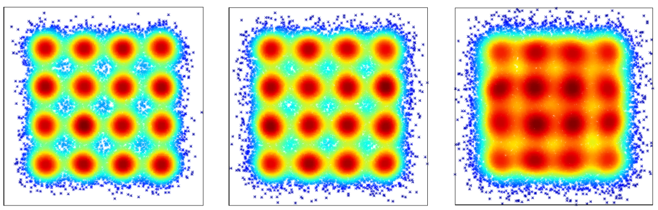


Fig. 3. Coherent optical communications with DNN-TEQ.



(a) Launch: -5 dBm (b) Launch: -3 dBm (c) Launch: -1 dBm

Fig. 4. Residual distortion of DP-16QAM after LE at 16-span NZDSF links.

listed below. DNN was introduced for optical signal-to-noise ratio (OSNR) monitoring in [4]. Modulation classification as well as OSNR monitoring was considered in [5], and a deep CNN showed an accurate performance in [6]. Deep learning-based network management and resource allocation were studied in [7], [8]. Analogously, traffic optimization with deep reinforcement learning was considered in [9], [10]. Various end-to-end learning methods, which jointly optimizes signal constellation and detection, have been proposed, e.g., [11]–[15], where a denoising auto-encoder is trained for nonlinear fiber channels. Also for receiver-end design, many DNN equalizers to compensate for fiber nonlinearity were introduced for coherent or non-coherent optical links, e.g., [16]–[21].

Note that large amounts of data necessary for deep learning are readily available in high-speed optical communications, where we can obtain terabits of data in a second [51]. In addition, the DNN can be massively parallelized in hardware implementation, which is suited for high-throughput communications. In deep learning, various techniques have been introduced, e.g., pre-training, rectified linear unit (ReLU), mini-batch, dropout, batch normalization, skip connection, inception, adaptive-momentum (Adam) stochastic gradient, adversarial, and long short-term memory architectures [3].

III. Deep Learning for Nonlinear Compensation

Similar to the other DNN equalizers, we focus on deep learning for fiber nonlinearity compensation. This paper has a distinguished contribution over existing literature as we propose a novel DNN-based TEQ suited for BICM-ID systems where state-of-the-art LDPC codes are employed.

A. Nonlinear Fiber-Optic Communications System

The optical communications system under consideration is depicted in Fig. 3. Three-channel dual-polarization (DP)-QAMs for 34 GBd baud and 37.4 GHz spacing 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 NZDSF has 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. The span loss is compensated by Erbium-doped fiber amplifiers (EDFA) with a noise figure of 5 dB, where total ASE noise is added just before the receiver. We use digital root-raised cosine filters with 10% rolloff at both transmitter and receiver. The receiver employs standard phase recovery and linear equalization (LE) to compensate for linear dispersion. Due to fiber nonlinearity, residual distortion after LE will limit the achievable information rates.

Fig. 4 shows a sample of residual distortion of DP-16QAM constellation after 31-tap least-squares LE for 16-span transmissions. We can observe that the constellation is more distorted with the increased launch power due to Kerr nonlinearity. To compensate for the residual nonlinear distortion, we introduce DNN-based TEQ, which exploits soft-decision feedback from the FEC decoder as shown in Fig. 3.

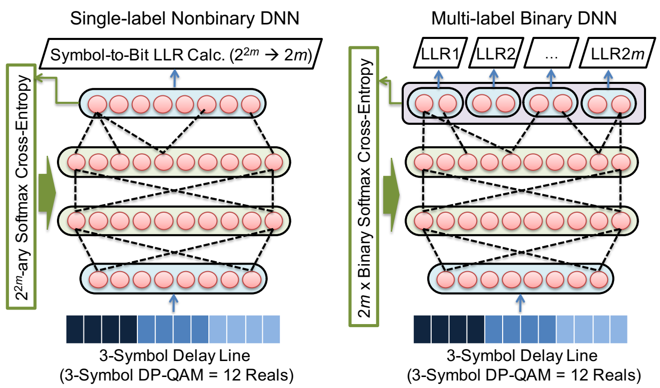
B. Scalable Deep Neural Network Equalization

Before introducing DNN-TEQ, we discuss the loss functions for training DNN equalizers suited for BICM. Consider DP-16QAM equalization, where 8 bits per symbol should be detected, leading to $2^8 = 256$ classes to identify. For such nonbinary classifications, we may use a single softmax classification shown in Fig. 5(a), like in [16]. However, this nonbinary (NB) DNN does not work well for higher-order DP-QAM in particular for a limited amount of training data. For example, DP-64QAM has 4096 classes to detect per symbol, which requires unrealistically huge training datasets.

To overcome the issues in high-order QAM, we shall use multi-label classification which employs multiple BCE losses as shown in Fig. 5(b). The binary DNN produces log-likelihood ratio (LLR), which can be directly fed into the SD-FEC decoder without external computation such as [16], [49]. This is a great advantage in practice because LLR calculation is cumbersome, especially for high-order modulation. Note that minimizing cross-entropy is equivalent to maximizing a lower bound of generalized mutual information.

C. Nonbinary vs. Binary DNN Equalization

We validate that DNN outperforms classical machine learning methods, specifically, quadratic discriminant analysis (QDA) and SVM (refer [20] for more comparisons). For multi-class SVM, we use the one-vs-one rule with a linear kernel as it worked best among several variants including one-vs-all and polynomial kernel. 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 of 80 km fiber links. It is seen that DNN offers the best performance among other methods, achieving greater than 1.2 dB gain over LE in highly nonlinear regimes. More importantly, the conventional DNN with nonbinary softmax cross-entropy does not perform well for high-order QAMs. It suggests that DNN equalizers using the BCE loss function has a great advantage not only for BICM compatibility but also for high-order QAM scalability.



(a) One 2^{2m}-ary softmax (b) $2m\times$ binary softmax

Fig. 5. Nonbinary and binary DNN equalizers for DP-2^mQAM.

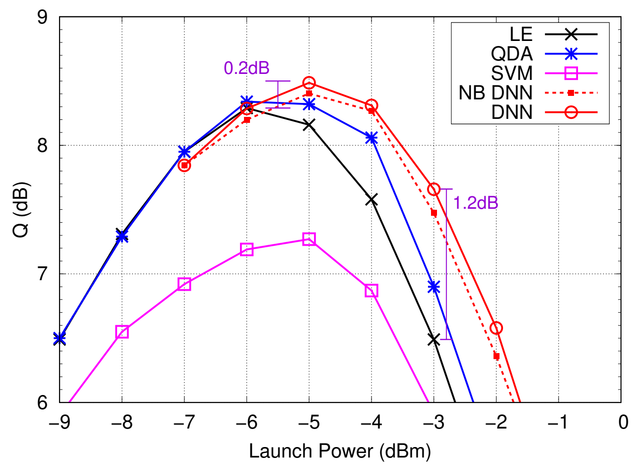


Fig. 6. Q factor comparisons for DP-4QAM 50-span NZDSF.

IV. Neural Turbo Equalization: DNN-TEQ

A. Nested Residual Network Architecture

Fig. 9 shows the architecture of our turbo DNN equalizer, which feeds distorted DP-QAM signals over consecutive $W = 3$-tap symbols to generate soft-decision LLR values for FEC decoding. The major extension from conventional DNN lies in the input layer which takes *a priori* (APR) information along with DP-QAM symbols. The APR side information comes from the FEC decoder representing intermediate soft-decision LLRs in run time. For efficient DNN training, the APR values having mutual information of $\mathcal{I}_{\text{in}}$ are synthetically generated via a Gaussian distribution following $\mathcal{N}((-1)^b\sigma^2/2, \sigma^2)$ where b is an original bit and $\sigma = J^{-1}(\mathcal{I}_{\text{in}})$ with $J^{-1}(\cdot)$ being ten Brink's J-inverse function [52], instead of considering a particular FEC decoder feedback.

The last layer has two branches, i.e., *extrinsic* (EXT) and *a posteriori* probability (APP) outputs, which uses a skip connection from the input layer to sum up EXT and APR. This nested residual network tries to train extrinsic message passing for TEQ realization. It was found that training a DNN model to minimize APP cross-entropy loss does not always minimize EXT cross-entropy loss accordingly, and vice versa.

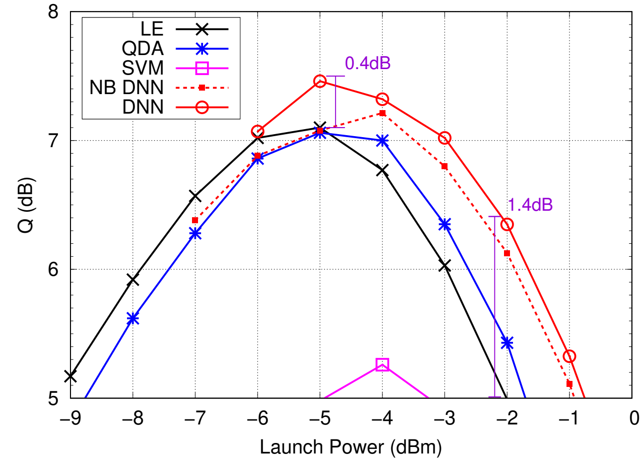


Fig. 7. Q factor comparisons for DP-16QAM 16-span NZDSF.

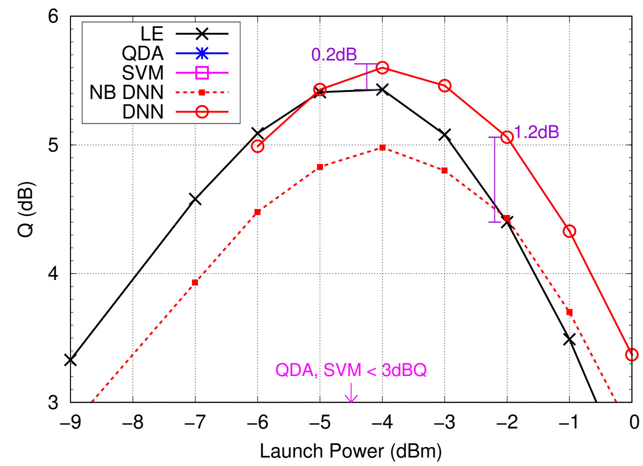


Fig. 8. Q factor comparisons for DP-64QAM 8-span NZDSF.

In order to keep both APP and EXT outputs reliable, we use a max-pooling layer following sigmoid cross-entropy loss.

The DNN uses four hidden layers, each of which consists of batch normalization, ReLU activation, and a fully-connected 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. Early stopping with a patience of 13 is employed for up to a maximum of 500 epochs.

B. EXIT Chart Analysis

Fig. 10 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 $\mathcal{I}_{\text{out}} = 1$ in the 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. 11, where

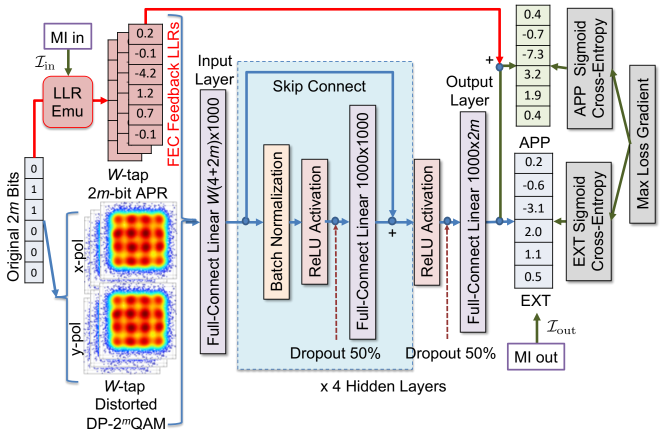


Fig. 9. DNN-TEQ architecture and min-max-loss training.

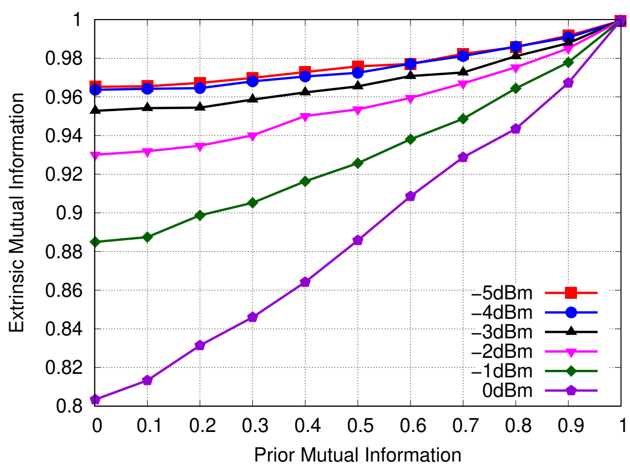


Fig. 10. EXIT chart of DNN-TEQ for DP-16QAM in 16-span DM links.

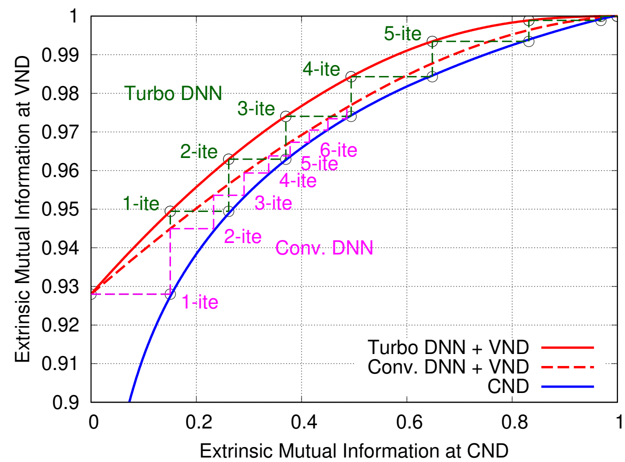


Fig. 11. Combined EXIT chart [52] of DNN-TEQ & LDPC decoder for DVB-S2 code rate $9/10$ (DP-16QAM in 16-span DM links at -2 dBm).

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 [52] 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 $\mathcal{I}_{\text{out}} = 1$. Whereas for DNN-TEQ, we can open up an EXIT tunnel between the VND and CND curves, that leads to a considerable acceleration of the decoder convergence to reach error-free condition within only a few iterations.

C. BER Performance

We assume the use of an outer Bose–Chaudhuri–Hocquenghem (BCH) $[30832, 30592]$ code with a rate of 0.9922 [51], having a minimum Hamming distance of 33. Based on the union (upper) bound, the bit-error rate (BER) threshold for this outer BCH code is at or above an input BER of 5×10^{-5} to achieve an output BER below 10^{-15}. Hence, a post-LDPC BER below 5×10^{-5} can be successfully decoded to a BER below 10^{-15} when this outer BCH code is used.

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 polynomial is $\rho(x) = x^{30}$. At a code rate of $5/6$, the variable and check degree polynomials are $\lambda(x) = \frac{2}{12}x^2 + \frac{9}{12}x^3 + \frac{1}{12}x^{13}$ and $\rho(x) = x^{22}$, respectively. We also consider an optimized degree distribution for DNN-TEQ as done analogously in [52], where the EXIT chart of DNN-TEQ in Fig. 10 is modeled with cubic functions and the EXIT curves of combined VND and DNN-TEQ are optimized for triple-degree check-concentrated distribution, which has two degrees of freedom to search for the best distribution. For example, the optimized LDPC code for a code rate of $5/6$ at a launch power of -4 dBm for DP-64QAM systems has a degree distribution of $\lambda(x) = 0.725x^2 + 0.25x^9 + 0.225x^{30}$.

Figs. 12 and 13 show the post-LDPC BER performance versus launch power of DP-16QAM and DP-64QAM, respectively, for 16, and 8 spans of NZDSF links. We used 4 and 8 turbo iterations respectively for DP-16QAM and DP-64QAM, where only one inner iteration between VND and CND for belief-propagation (BP) decoding is performed for each outer turbo iteration. We compare DVB-S2 LDPC codes for LE, DNN and DNN-TEQ and our optimized LDPC code for DNN-TEQ. From the figures, we can observe the following results:

- Although DNN nonlinear compensation can improve BER performance of LE, achieving a BER of BCH threshold is mostly in failure.
- DNN-TEQ can significantly improve the BER performance of DNN to reach the threshold and about ± 2 dB margin around an optimal launch power is realized.
- Optimizing LDPC codes for DNN-TEQ can offer an additional marginal improvement over the standard DVB-S2 LDPC codes for the whole range of launch power.

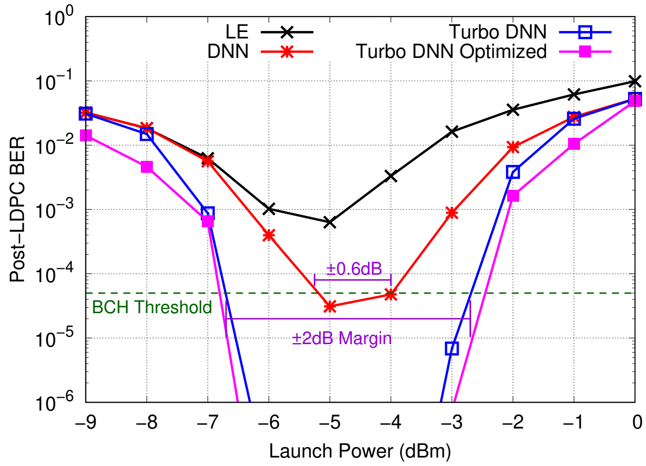


Fig. 12. BER performance for DP-16QAM 16-span NZDSF DM links (LDPC code rate $9/10$, 4-iteration BP decoding).

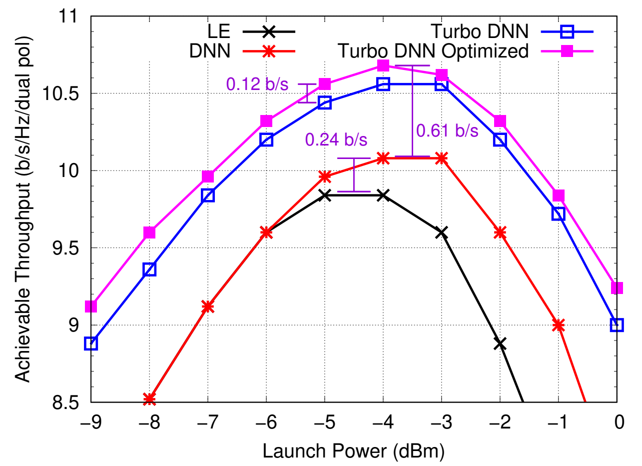


Fig. 14. Achievable rate for DP-64QAM 8-span NZDSF DM links (variable-rate LDPC codes, 8-iteration BP decoding).

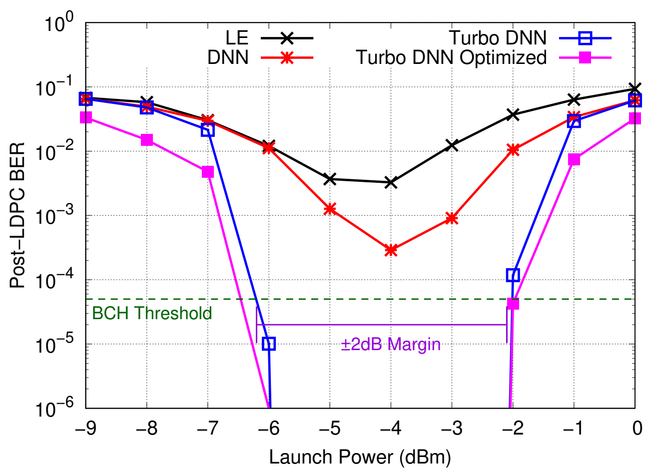


Fig. 13. BER performance for DP-64QAM 8-span NZDSF DM links (LDPC code rate $5/6$, 8-iteration BP decoding).

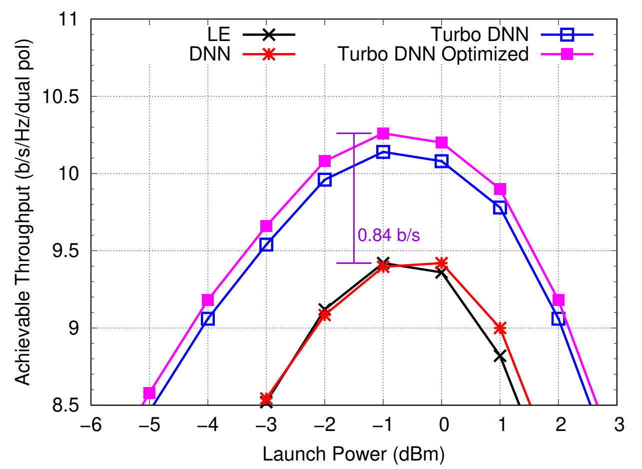


Fig. 15. Achievable rate for DP-64QAM 20-span SSMF UM links (variable-rate LDPC codes, 8-iteration BP decoding).

D. Achievable Rate Performance

The BER improvement with our proposed DNN-TEQ implies that we can increase the achievable throughput when the code rate is adaptively optimized. Fig. 14 shows achievable rate performance for DP-64QAM at 8-span NZDSF links. Here, we use the same variable node degree of DVB-S2 rate $5/6$ and plot the largest code rate such that the post-LDPC BER meets the BCH threshold by varying the check node degree to be a target rate. From this figure, we can see that the DNN nonlinear compensation can improve the performance of LE by 0.7 b/s/Hz in the nonlinear regimes, and the achieved gain in the peak throughput is about 0.24 b/s/Hz. Our DNN-TEQ offers a remarkable BICM-ID gain over the whole range of launch powers, achieving a throughput improvement of 0.61 b/s/Hz over the DNN when the LDPC code is optimized. A total throughput improvement of 0.85 b/s/Hz from the standard LE was achieved by the proposed DNN-TEQ.

It should be noted that our DNN-TEQ achieved a remarkable gain over the DNN even in the linear regimes. The gain in the linear regimes is purely due to the rate improvement from BICM to BICM-ID, whose information-theoretic bound approaches the Shannon limit at lower SNRs. It does not contradict the results in Fig. 13, where no gain was observed at very low (and high) launch power cases. For Fig. 13, we used a rate-$5/6$ DVB-S2 code, which was not decodable at low launch powers, resulting in useless soft-decision feedback having nearly zero mutual information. With a proper choice of code rates, TEQ can take advantage of FEC feedback information to boost the performance. We also note that the achievable rate evaluation for all cases in Fig. 14 includes loss caused by the practical constraint of finite 8-iteration BP decoding.

We finally evaluate the DNN-TEQ in dispersion unmanaged (UM) links in Fig. 15, which shows the achievable rates for 20-span UM standard single-mode fiber (SSMF) links ($D = 17$ ps/nk/km, $\gamma = 1.2$ /W/km). Note that we extended the reach from 8 spans to 20 spans for UM links since the nonlinear distortion per span is relatively weak. It was confirmed that the DNN-TEQ still offers a significant rate improvement even for the UM links. However, the improvement was mostly due to BICM-ID gain and there was no additional gain due to nonlinearity compensation at higher launch powers. This is because the considered 3-tap DNN could not handle a long channel memory in the SSMF UM links having approximately 200-times longer memory than 8-span NZDSF DM links. Engineering the neural network architectures suited for tackling long-memory fiber nonlinearity still remains a challenge.

V. Conclusions

We extended DNN machine learning techniques to TEQ for improved nonlinear compensation in coherent fiber communications. We first verified that DNN trained with binary cross-entropy loss can outperform various machine learning techniques to compensate for fiber nonlinearity. Through EXIT chart analysis, we then confirmed that the proposed DNN-TEQ offers decoder acceleration by feeding intermediate soft-decision LLR from the LDPC decoder. Our DNN-TEQ significantly improves BER performance through the turbo iteration. We also investigated LDPC code design based on the EXIT chart of DNN-TEQ, and demonstrated that the proposed DNN-TEQ with optimized LDPC codes can improve the achievable throughput by 0.85 b/s/Hz over linear equalization with standard LDPC codes. To the best of our knowledge, this is the first paper investigating TEQ based on DNN for fiber nonlinearity mitigation. Further improvement by dealing with long nonlinear channel memory, particularly in dispersion unmanaged links, remains as future work.

References

[1] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, July 2006.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[3] S. Hockreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[4] T. Tanimura, T. Hoshida, T. Kato, S. Watanabe, J. C. Rasmussen, M. Suzuki, and H. Morikawa, "Deep learning based OSNR monitoring independent of modulation format, symbol rate and chromatic dispersion," *European Conference on Optical Communication (ECOC)* Sep. 2016.

[5] F. N. Khan, K. Zhong, X. Zhou, W. H. Al-Arashi, C. Yu, C. Lu, and A. P. Lau, "Joint OSNR monitoring and modulation format identification in digital coherent receivers using deep neural networks," *Opt. Expr.* vol. 25, no. 15, pp. 17767-17776, July 2017.

[6] D. Wang, M. Zhang, J. Li, Z. Li, J. Li, C. Song, and X. Chen, "Intelligent constellation diagram analyzer using convolutional neural network-based deep learning," *Opt. Expr.*, vol. 25, no. 15, pp. 17150-17166, July 2017.

[7] J. Guo and Z. Zhu, "When deep learning meets inter-datacenter optical network management: Advantages and vulnerabilities," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4761–4773, Oct. 2018.

[8] A. Yu, H. Yang, W. Bai, L. He, H. Xiao, and J. Zhang, "Leveraging deep learning to achieve efficient resource allocation with traffic evaluation in datacenter optical networks," *Optical Fiber Commun. Conf. (OFC)*, Mar. 2018.

[9] X. Luo, C. Shi, L. Wang, X. Chen, Y. Li, and T. Yang, "Leveraging double-agent-based deep reinforcement learning to global optimization of elastic optical networks with enhanced survivability," *Opt. Expr.*, vol. 27, no. 6, pp. 7896–7911, Mar. 2019.

[10] Y. Tang, H. Guo, T. Yuan, X. Gao, X. Hong, Y. Li, J. Qiu, Y. Zuo, and J. Wu, "Flow Splitter: A deep reinforcement learning-based flow scheduler for hybrid optical-electrical data center network," *IEEE Access*, vol. 7, pp. 129955–65, Sep. 2019.

[11] 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," *European Conf. Optical Commun. (ECOC)*, Sep. 2018.

[12] 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," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–55, Aug. 2018.

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

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

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

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

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

[18] 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," *Optical Fiber Commun. Conf. (OFC)*, p. Th4D-5, Mar. 2018.

[19] 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," *Optical Fiber Commun. Conf. (OFC)*, p. M2B.2, Mar. 2018.

[20] 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*, p. SpM4G.1, July 2018.

[21] T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization to mitigate fiber nonlinearity," *European Conf. Optical Commun. (ECOC)*, p. Tu.1.B.1, Sep. 2019.

[22] A. D. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Aug. 2009.

[23] M. Secondini, E. Forestieri, and G. Prati, "Achievable information rate in nonlinear WDM fiber-optic systems with arbitrary modulation formats and dispersion maps," *J. Lightw. Technol.*, vol. 31, no. 23, pp. 3839–3852, Dec. 2013.

[24] J. Renaudier, G. Charlet, P. Tran, M. Salsi, and S. Bigo, "A performance comparison of differential and coherent detections over ultra long haul transmission of 10Gb/s BPSK," *Optical Fiber Commun. Conf. (OFC)*, p. OWM1, Mar. 2007.

[25] N. Alić, G. C. Papen, R. E. Saperstein, L. B. Milstein, and Y. Fainman, "Signal statistics and maximum likelihood sequence estimation in intensity modulated fiber optic links containing a single optical preamplifier," *Opt. Express*, vol. 13, no. 12, pp. 4568–4579, June 2005.

[26] Y. Cai, D. G. Foursa, C. R. Davidson, J. X. Cai, O. Sinkin, M. Nissov, and A. Pilipetskii, "Experimental demonstration of coherent MAP detection for nonlinearity mitigation in long-haul transmissions," *Optical Fiber Commun. Conf. (OFC)*, p. OTuE1, Mar. 2010.

[27] 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.*, vol. 20, no. 14, pp. 15769–15780, July 2012.

[28] 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," *IEEE JSAC*, vol. 26, no. 6, pp. 73–83, Aug. 2008.

[29] 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*, vol. 2, no. 4, pp. 593–599, Aug. 2010.

[30] 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," *Optical Fiber Commun. Conf. (OFC)*, p. JW2A.59, Mar. 2012.

[31] K. V. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *J. Lightw. Technol.*, vol. 15, no. 12, pp. 2232–2241, Dec. 1997.

[32] F. P. Guiomar, J. D. Reis, A. Teixeira, and A. N. Pinto, "Mitigation of intra-channel nonlinearities using a frequency-domain Volterra series equalizer," *European Conf. Optical Commun. (ECOC)*, p. Tu.6.B.1, Sep. 2011.

[33] 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," *European Conf. Optical Commun. (ECOC)*, p. We3C6, Sep. 2013.

[34] X. Li, X. Chen, G. Goldfarb, E. Mateo, I. Kim, F. Yaman, and G. Li, "Electronic post-compensation of WDM transmission impairments using coherent detection and digital signal processing," *Opt. Express*, vol. 16, no. 2, pp. 880–888, Jan. 2008.

[35] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.

[36] E. Ip, N. Bai, and T. Wang, "Complexity versus performance tradeoff in fiber nonlinearity compensation using frequency-shaped, multi-subband backpropagation," *Optical Fiber Commun. Conf. (OFC)*, p. OThF4, Mar. 2011.

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

[38] N. Irukulapati, H. Wymeersch, and P. Johannisson, "Extending digital backpropagation to account for noise," *European Conf. Optical Commun. (ECOC)*, p. We.3.C.4, Sep. 2013.

[39] X. Xie, F. Yaman, X. Zhou, and G. Li, "Polarization demultiplexing by independent component analysis," *IEEE Photon. Technol. Lett.*, vol. 22, no. 11, pp. 805–807, June 2010.

[40] 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 Photon. Technol. Lett.*, vol. 27, no. 4, pp. 387–390, Dec. 2014.

[41] 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.*, vol. 40, no. 21, pp. 5113–5116, Nov. 2015.

[42] H. Zhao and J. Zhang, "Adaptively combined FIR and functional link artificial neural network equalizer for nonlinear communication channel," *IEEE Trans. Neural Networks*, vol. 20, no. 4, pp. 665–674, Feb. 2009.

[43] D. J. Sebald and J. A. Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Processing*, vol. 48, no. 11, pp. 3217–3217, Nov. 2000.

[44] 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," *Optical Fiber Commun. Conf. (OFC)*, p. Th2A-49, Mar. 2016.

[45] K.-P. Ho and J. M. Kahn, "Electronic compensation technique to mitigate nonlinear phase noise," *J. Lightw. Technol.*, vol. 22, no. 3, pp. 779–783, Mar. 2004.

[46] S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. Neural Networks*, vol. 4, no. 4, pp. 570–590, July 1993.

[47] 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," *Optical Fiber Commun. Conf. (OFC)*, p. M3A-3, Mar. 2014.

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

[49] T. Yoshida, K. Matsuda, K. Kojima, H. Miura, K. Dohi, M. Pajovic, T. Koike-Akino, D. S. Millar, K. Parsons, and T. Sugihara, "Hardware-efficient precise and flexible soft-demapping for multi-dimensional complementary APSK signals," *European Conf. Optical Commun. (ECOC)*, p. Th.2.P2.SC3.27, Sep. 2016.

[50] C. Häer and H. D. Pfister, "Nonlinear interference mitigation via deep neural networks," *Optical Fiber Commun. Conf. (OFC)*, p. W3A.4, Mar. 2018.

[51] D. S. Millar, R. Maher, D. Lavery, T. Koike-Akino, M. Pajovic, A. Alvarado, M. Paskov, K. Kojima, K. Parsons, B. C. Thomsen, S. J. Savory, and P. Bayvel. "Design of a 1 Tb/s superchannel coherent receiver," *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1453–1463, Mar. 2016.

[52] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE TCOM*, vol. 52, no. 4, pp. 670–678, May 2004.

Toshiaki Koike-Akino (M'05–SM'11) received the B.S. degree in electrical and electronics engineering, M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010 he was a Postdoctoral Researcher at Harvard University, and joined MERL, Cambridge, MA, USA, in 2010. His research interests include signal processing for data communications and sensing. He received the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award.

Ye Wang (S'06–M'12–SM'19) received the B.S. degree in Electrical and Computer Engineering from Worcester Polytechnic Institute, Worcester, MA, in 2005, and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from Boston University, Boston, MA, in 2009 and 2011, respectively. In 2012, he joined MERL, Cambridge, MA, where he is now a Principle Research Scientist. He has also contributed to standardization activity in the INCITS-M1 US National Body for Biometrics Standards and the ISO/IEC-JTC1/SC37 Subcommittee on Biometrics, where he served as a project editor. Prior to joining MERL, he was a postdoctoral researcher in the Department of Electrical and Computer Engineering at Boston University, and had worked as a Firmware Engineer at AgaMatrix, Salem, NH. His technical interests include information theory, machine learning, signal processing, communication, and data privacy/security.

David S. Millar (S'07–M'11) was born in Manchester, U.K., in 1982. He received the M.Eng. degree in electronic and communications engineering from the University of Nottingham, Nottingham, U.K. in 2007, and the Ph.D. degree in optical communications from UCL, London, U.K., in 2011. He is currently working at MERL, Cambridge, MA, USA. His research interests include coherent optical transmission systems, digital coherent receiver design, coded modulation for optical communications, and digital nonlinearity mitigation.

Keisuke Kojima (S'82–M'84–SM'13) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1981, 1983, and 1990, respectively. He also received the M.S. degree from the University of California, Berkeley, CA, USA, in 1982. He worked for eight years at the Central Research Laboratory, Mitsubishi Electric Corp., from 1983 on the research of narrow linewidth DFB and DBR lasers, grating-coupled surface-emitting-lasers, and optical neural networks. He spent nine years at AT&T/Lucent Bell Laboratories on the R&D of uncooled Fabry–Perot and DFB lasers, vertical-cavity surface-emitting lasers, and passive optical network systems, first as a Member of Technical Staff, and later as a Technical Manager. He also worked at Agere Systems, Denselight Semiconductors, and TriQuint Semiconductors on optical devices and modules, and optical systems testbed. He has been with MERL, Cambridge, MA, USA, since 2005, where he is currently working on the design of photonic-integrated circuits and coherent optical systems using deep learning, as a Senior Principal Research Scientist. He has more than 200 publications in journals and conference proceedings. He is a Fellow of the Optical Society of America, and Associate Editor of IEEE/OSA Journal of Lightwave Technology.

Kieran Parsons (M'07–SM'09) received the B.Eng. and Ph.D. degrees in electronic and communications engineering from the University of Bristol, Bristol, U.K., in 1992 and 1996, respectively. During 1997–2002, he was with Nortel Networks, Ottawa, Canada, where he worked on wireless and long-haul optical communications system architecture and design. From 2004–2006, he worked on carrier-grade mesh WiFi RF system design at BelAir Networks (now part of Ericsson) and from 2006–2009 on 10-G PHY device development with Applied Micro, both in Kanata, Canada. In 2009, he joined MERL, Cambridge, MA, USA, where he is currently Team Leader and Senior Principal Research Scientist. His research interests include digital signal processing and coherent optical transmission systems.