Coupled Transceivers-Fiber Nonlinearity Compensation Based on Machine Learning for Probabilistic Shaping System

Tu Thanh Nguyen, Tingting Zhang, Elias Giacoumidis, Paul Harper and Andrew Ellis

Abstract—In this paper, we have studied the impact of probabilistic constellation shaping (PCS) signals on transceivers and proposed, for the first time, an artificial neural network (ANN)-based nonlinearity conpensation (NLC) to compensate for coupled-nonlinear distortion from the transceivers and fiber propagation (Kerr effects) for a PCS system. A PCS dual-polarization 28 GBaud system equipped with an ANN-based nonlinear compensator is experimentally demonstrated. The performance of the proposed ANN-based NLC is first studied for transceivers nonlinearity compensation in the PCS 64/256-QAM system and its performance improvement is also compared to that of the uniformly constellation 64/256-QAM formats. We then investigate experimentally the effectiveness of the proposed scheme to compensate for coupled-nonlinearity (transceivers and fiber-induced nonlinearity) for uniform and shaped 256-QAM signals over a fiber channel of 300 km and 500 km inline Erbium-doped-amplification link. The Experimental results show that a SNR performance gain up to 1 dB can be achieved for compensating the transceivers nonlinearity with the artificial neural network (ANN)-based NLC. In the present of the coupled transceivers and fiber-induced nonlinearity, a mutual information enhancement of ∼0.25 bits/symbol is demonstrate experimentally for a fiber link of up to 500 km with the aid of the proposed NLC.

Keywords—*Transceiver nonlinearity, machine learning, ANN, nonlinear equalizer, probabilistic shaping, fiber nonlinearity.*

I. Introduction

To meet the fast-increasing demand of data traffic, high-order quadrature amplitude modulation (QAM) formats combined with probabilistic constellation shaping (PCS) have attracted a lot of attention in recent years. This combination enables both high spectral efficiency (SE) and flexible transmissions [1]–[5]. Adaptable transmission rates can be easily realized by adjusting only the shaping factor in stead of modifying forward-error-correction (FEC), changing systems bandwidth, using hybrid modulation format and/or sub-carrier multiplexing. The idea of PCS is to shape the signal constellation as close as possible to the optimum constellation for the channel. As a proof of concept, a near-optimal signal-to-noise ratio (SNR) gain of 1.53 dB is feasible when employing the probabilistic shaping technique in a Gaussian channel [2]. Although the PCS is one of the most promising candidates for next generation transponders, its impacts on the digital signal processing (DSP) chain and on the transceivers including fiber channel have not been fully explored yet and thus, it requires further investigation.

Generally speaking, the implementation of high-order modulation formats such as 64-QAM and beyond is often a big challenge due to the requirement for high SNR, high effective number of quantization-bits of digital-to-analog converters (DACs)/analog-to-digital converters (ADCs). This problem is envisaged that will be more severe when PCS signals are modulated. In addition, PCS signals may require DSP adaptation for data recovery because conventional DSP algorithms such as blind phase noise recovery and linear equalizer are generally not compatible with shaping systems [1], [6], [7].

Due to their higher peak-to-average power ratio (PAPR) than that of the uniformly distributed constellations, the PCS signals may require a better transceivers linearity, which is yet is practically limited by the imperfection of transceiver devices such as DAC/ADC, power amplifiers and optical modulators. In literature, the nonlinearity transceivers for uniform constellation systems can be compensated using digital filters [8], [9]. Nevertheless, it is difficult to estimate the exact coefficients of these filters as the result of the nonlinear mixing from different devices of the transceivers, especially in a mesh optical network. To partly deal with this problem, a supervised machine-learning-based technique, namely ANN, has been recently applied numerically for uniform 64-QAM as a pre-distortion compensation [10]. However, in [10] authors considered low resolution DAC at the transmitter and ignored nonlinear contributions from other components such as optical modulators, ADCs and fiber-induced nonlinearity. Since the increased PAPR places greater demand on effective number-of-bits (ENoB) of DACs/ADC and increases sensitivity to transponders and fiber-induced nonlinearity, more nonlinear distortion is expected when deploying PCS signals on the same system infrastructure. The penalty due to the coupled nonlinearity may be considerable in practical PCS systems. This coupled nonlinearity, however, is complicated to describe analytically due to the interplay between the nonlinearity,

Tu Thanh Nguyen, Tingting Zhang, Paul Harper and Andrew Ellis are with the Aston Institute of Photonic Technologies, Aston University, Aston St, Birmingham B4 7ET, UK. Emails: t.nguyen14@aston.ac.uk (corresponding), t.zhang32@aston.ac.uk, p.harper@aston.ac.uk and andrew.ellis@aston.ac.uk

Elias Giacoumidis was with the Radio and Optical Communications Laboratory, School of Electronic Engineering, Dublin City University, Glasnevin, Dublin 9, Ireland & SFI CONNECT Research Centre; and now is with VPIphotonics GmbH Carnotstraße 6, 10587 Berlin, Germany (email: ilias.giakoumidis@vpiphotonics.com).

bandwidth, and system memory (dispersion) of multiple components of the link, which not only vary from link to link, but which may also vary with time.

In order to compensate for the transceivers nonlinearity, we have recently proposed to use the ANN algorithm at the receiver for PCS systems as the post-compensation [11]. In this paper, we extend our work and experimentally demonstrate the effectiveness of the proposed scheme for compensating not only the nonlinear distortion from transceivers but also the Kerr nonlinearity in an optical channel of up to 500 km standard single mode fiber (SSMF). Initially, the proposed approach for transceivers nonlinearity compensation is experimentally demonstrated for the PCS dual-polarization (DP) 28 GBaud 64/256-QAM system under different shaping factors. The performance the proposed scheme for compensating the coupled transceivers-fiber nonlinearity is then studied experimentally for uniform and shaped 256-QAM signals over an optical link of 300 km and 500 km inline Erbium-doped fiber amplifier (EDFA)-based SSMF. Different ANN configurations for the nonlinear compensation (NLC) are also discussed under the fiber transmission. While there is no significant improvement observed in PCS 64-QAM, a SNR gain of 1 dB is experimentally demonstrated for combating the transceivers nonlinearity in PCS 256-QAM using the proposed ANN-based NLC scheme. On the other hand, a mutual information (MI) enhancement of ∼0.25 bits/symbol with the proposed NLC for compensating the coupled transceivers-fiber nonlinearity is demonstrated experimentally for the PCS 256-QAM system over a transmission distance of 500 km SSMF.

II. FUNDAMENTALS OF PROBABILISTIC SHAPING AND PRINCIPLE OF ANN-BASED NONLINEAR EQUALIZER

A. Probabilistic shaping: a brief introduction

The basic principle of PCS is to code the signal such that different a-priori probabilities belong to different constellations. Generally, lower-energy symbols (near the origin) are taken place more often than higher-energy signals (far from the origin) after PCS. This results in a reduction in average transmitted power, and therefore higher fiber nonlinearity tolerance. In this work, we deploy the well-known Maxwell-Boltzmann (MB) distribution for generating a set of probability mass functions (PMFs), $P_X(x_j)$, for each modulation format as [2]:

$$P_X(x_j) = \frac{e^{-\kappa|x_j|^2}}{\sum_{k=1}^{M} e^{-\kappa|x_k|^2}}, \quad j = 1, 2, ..., M, \quad (1)$$

where x_j is one of the input alphabet, M is the modulation size (e.g. $M = 64$ for 64-QAM) and κ is the shaping factor. For maximum transmission capacity, the shaping factor needs to be optimized. Generally, this optimization is a function of signal power, modulation format and SNR [2]. For each complex QAM, two-dimension (2D) optimization (representing the real and imaginary parts) needs to be carried out. To simplify this process, square QAM is deployed and this is the reason 128-QAM is not considered here. The 2D optimization for PCS QAM can be simplified to one-dimensional (1D) optimization for the corresponding PCS pulse amplitude modulation
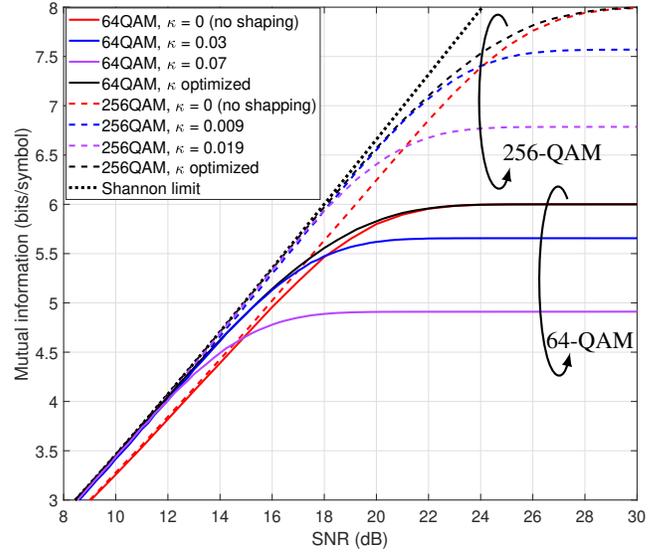
Fig. 1: Comparison in terms of mutual information between 64-QAM and 256-QAM under different shaping factors. Note that $\kappa = 0$ means uniform constellation (no shaping).

(PAM). In this work, we chose 64-QAM and 256-QAM, for the experimental verification. We have also assumed that the location of shaped symbols are similar to the uniformly distributed constellation, i.e. their real or imaginary amplitudes are integers of $\pm(2k+1)$, $k = 0, 1, ..., \sqrt{M}/2 - 1$.

Fig. 1 shows the comparison in terms of MI (representing the number of bits per symbol, also known as symbol-wise achievable information rate), between 64-QAM and 256-QAM under different shaping factors. Note that $\kappa = 0$ corresponds to the uniform constellation (no shaping). The MI is estimated via Monte-Carlo simulation of N input-output symbol pairs (x_k, y_k) as

$$\text{MI} = \frac{1}{N} \sum_{k=1}^{N} \log_2 \frac{q_{Y|X}(y_k|x_k)}{\sum_{x_j=1}^{M} q_{Y|X}(y_k|x_j) P_X(x_j)}, \quad (2)$$

where $q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{|y-x|^2}{2\sigma^2})$ is the Gaussian channel transition probability density function with σ^2 and (x, y) being the noise variance and the input-output pair of the channel, respectively. In this figure, the Shannon capacity limit in Gaussian channel is also given as $\log_2(1 + \text{SNR})$ for reference (labeled as "Shannon limit").

For 64-QAM and 256-QAM, four shaping rates are considered to study in this paper: $\kappa_1 = 0.03$ and $\kappa_2 = 0.07$ for PCS 64-QAM; and $\kappa_3 = 0.009$ and $\kappa_4 = 0.019$ for PCS 256-QAM. Fig. 1 shows the MI evolution of 64/256-QAM under different shaping rates with respect to SNR. For a single modulation format, the MI curves associated with uniform constellation (no shaping) and optimal constellation (shaping rate optimized at each SNR for a full shaping gain) are also provided. The chosen shaping rates are adopted from [12] in which only two fixed PMFs are sufficient for a wide
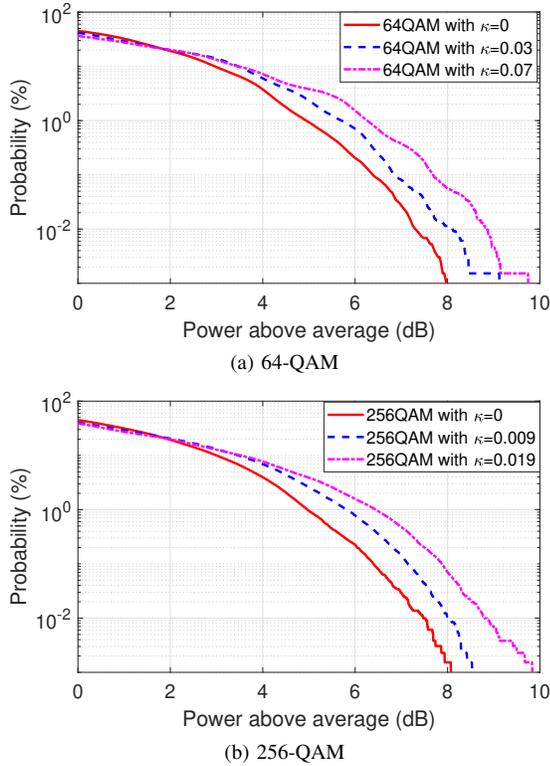
Fig. 2: PAPR comparison of QAM signals under different shaping rates: (a) 64-QAM and (b) 256-QAM.

SNR range with a negligible penalty (at about 0.1 dB of SNR) to the optimum shaping. The entropy, which indicates the maximum information rate at infinite SNR, of the investigated PCS systems with shaping rates of κ_1, κ_2, κ_3 and κ_4 is 5.66, 4.91, 7.57 and 6.79, respectively. For fiber channel, these shaping factors are also good choices as they give a relative maximum shaping gain and thus, further optimization is not necessary to carry out [12].

Fig. 1 also indicates there is a SNR range for which constellation shaping should be applied for each modulation format. For instance, in order to maintain a noticeable shaping gain, the PCS should be applied for systems under SNR ranges <21 dB and <27 dB for 64-QAM and 256-QAM, respectively. Taking into account the constraint of less than 0.1 dB SNR penalty to the full shaping gain using the aforementioned rates and potential higher implementation penalty of 256-QAM versus 64-QAM, the suggested SNR ranges for $\kappa_{1,2,3,4}$ are roughly [8 dB, 12 dB], [12 dB, 16 dB], [14 dB, 18 dB] and [18 dB, 22 dB], respectively.

B. Impact of PCS signals on transceivers

It is clear that PCS changes the statistical property of the transmitted signal. In this section, the figure of PAPR metric is investigated for PCS signals with different shaping factors. The PAPR of a signal $x(n)$ is calculated as PAPR (dB) $= 10\log_{10}\frac{\max|x(n)|^2}{\mathbb{E}[|x(n)|^2]}$, where $\mathbb{E}\{.\}$ is the expectation
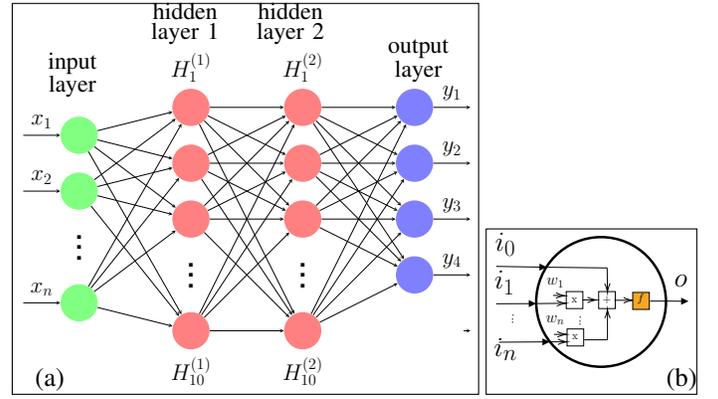
Fig. 3: (a) Structure of artificial-neural-network-based nonlinear compensation (ANN-based NLC). (b) Input-output relationship of a "neuron" in the hidden and output layer.

operator. Fig. 2 shows the PAPR comparison in terms of complementary cumulative distribution function (CCDF) of QAM signals with and without PCS. The vertical axis is the CCDF showing how often a certain PAPR value in the horizontal axis is exceeded. The CCDF were measured after pulse-shape filtering using a root-raised-cosine (RRC) filter (0.1 roll-off factor and up-sampling at 2 samples per symbol). For comparison purpose, two aforementioned shaping factors in the Section 2 were considered for each modulation format. As shown from Fig. 2-a and Fig. 2-b, the trend of the PAPR curves for the two studied QAMs is almost identical. The shaping signals exhibit larger PAPR than the uniform constellation cases ($\kappa = 0$) and the more shaping, the worse PAPR. Specifically, for 64-QAM, at the same probability of 1 %, the PAPR increases by 0.7 dB and 1.5 dB when the shaping factor increases from 0 to 0.03 and from 0 to 0.07, respectively. The same amount of PAPR increment is also seen for 256-QAM signals. The PAPR increments result in the increase nonlinear distortions, unless the linear operation ranges are increased accordingly. The higher PAPR also imposes more Kerr-induced nonlinear noise which is proportional to the instantaneous power of the signal propagating through the optical fiber.

C. Principle of ANN-based NLC

Machine learning approaches for fiber communications have gained a lot of attention and been intensively studied in optical communications recently due to their capability of providing practical solutions for complicated systems in which analytic approach is not available. Several common applications of the machine learning in optical communications can be performance monitoring and fiber-induced nonlinearity compensation [13]–[18]. Hitherto machine learning algorithms have been harnessed mostly for uniform QAMs. Very recently, however, a few machine learning-based algorithms such as ANN and deep-learning algorithms have been implemented for probabilistic shaping systems but mainly for the purpose of constellation optimization [19], [20] .

The proposed ANN-based NLC is one of the regression-based supervised-learning algorithm. Fig. 3-a shows the structure of a simple feed-forward network deployed in this paper as the NLC. This structure is a typical configuration of ANN. In short, it comprises of n-input x 4-output with two hidden-layers each of 10 nodes (also known as "neurons"). The number of inputs will be a function of memory depth m as $4(m+1)$ whereas the factor of 4 represents real and imaginary components of dual-polarization signals. The aim of ANN is to find a function that maps the input to the desired target through a number of intermediate steps produced by the neurons in the network.

Within each neuron (Fig. 3-b), there are three calculations taking place: (1) weight multiplication of inputs (subscripts $1, 2, ..., n$), (2) adding a bias (the input with subscript 0) to the weighted data, and (3) passing the results of (2) through an activation function. In the proposed ANN-based NLC scheme, the activation function used in each hidden layer is a nonlinear hyperbolic tangent sigmoid transfer function, whereas 4 neurons of the output layer use a linear transfer function.

The ANN-based NLC was operated in two phases: the training and operational phase. In a supervised learning scheme, the training phase needs information of transmitted data and optimize weights of the ANN is figured out via the batch Levenberg-Marquardt back-propagation algorithm (using gradient descent) [21]. Around 2^{15} transmitted symbols are used in the training phase in which the ratios of 70 % (batch size), 15 % and 15 % are dedicated for the training, validation and testing, respectively. At the very begining of the training, initialize weights, $\boldsymbol{\theta}_0$, are generated randomly. For the kth batch (epoch), the model output, $\boldsymbol{y}_{\theta_k}(\boldsymbol{x})$ is calculated by implementing the forward propagation of the input $\boldsymbol{x}$. Then, the difference between targets and model outputs, also known as loss function $J_\theta = \mathbb{E}\{\boldsymbol{y}_{\theta_k}(\boldsymbol{x}) - \boldsymbol{x}\}^2$, is computed. Next, the gradients of the loss function, i.e. partial derivatives of the loss function with respect to all the weights $\frac{\partial J_\theta}{\partial \theta_{lm}}$, is calculated in a back-propagation manner. The weight parameters are then updated in response to the gradients for the next epoch as $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \frac{\partial J_\theta}{\partial \theta}$. This cycle repeats in the $(k+1)^{\text{th}}$ epoch until the minima of the loss function or the maximum number of epochs is reached. The number of epochs is 100. After the training, the inverted nonlinear function with the optimum parameter set, reflects the coupled nonlinearity of the transceivers and from the transmission fiber, is used as the NLC.

In the operational phase, the received signals are simply compensated by passing the received signals through the trained ANN. In optical back-to-back configuration, we assumed that the nonlinear distortion from transceivers was static or time-slowly varying, and therefore the training phase was only performed once at the optimum condition (i.e. highest SNR). This training process may be repeated periodically if necessary (during initialization/calibration stages, for example). Regarding fiber channel, however, the ANN re-training is required. This depends on how long the link is and also how deeply shaped signal is. This will be discussed more in detail in Section IV.

Regarding the complexity of the proposed scheme, typically the number of real multiplications can be used as the figure of merit because the complexity of other operations such as additions take a small proportion of the total complexity. Let denote the number of nodes in the kth hidden layer being N_k^{h}. The number of real multiplications per symbol of the ANN-based NLC operating in a forward manner is easily calculated as $4(m + 1)N_1^{\text{h}} + N_1^{\text{h}}N_2^{\text{h}} + N_2^{\text{h}}N_3^{\text{h}} + ... + N_n^{\text{h}}4$. Note that the activation function in each node can be implemented efficiently by using a look-up-table [22]. Thus, the ANN-based NLC requires only 380 real multiplications per symbol ($m = 5, N_1^{\text{h}} = 10, N_2^{\text{h}} = 10$).

III. EXPERIMENTAL SETUP

The experimental setup of the 28 GBaud PCS DP 64/256-QAM system is shown in Fig. 4. At the transmitter, four streams of 8/16-PAM data, each of ∼60 000 symbols, with desired PMFs according to the aforementioned shaping rates were generated randomly. Then, PCS 64/256-QAM signals on each polarization were formed by combining two independently shaped 8/16-PAM sequences which represent their real and imaginary components. The shaped symbols were normalized for a unit average power and multiplexed with 5 % of 4-QAM pilot symbols (i.e. 1 pilot in every 20 symbols) to aid the DSP algorithms at the receiver for channel equalization and phase noise compensation. The power of 4-QAM pilot symbols was also normalized to 1 before the multiplexing. There was no DSP adaptation for the implemented PCS in this paper as they were replied on pilot-aided symbols. A RRC filter with a roll-off factor of 0.1 was then applied and up-sampled at 2 samples-per-symbol. After this off-line processing, the signal was loaded into the Keysight M8195A arbitrary waveform generator (4-channel 8-bit DAC sampling at 56 GSa/s) and subsequently converted into the optical domain by using a conventional DP optical modulator (laser linewidth ∼100 kHz on 192.4 THz) - the Tektronix OM5110. An EDFA followed by a variable optical attenuator (VOA) was used to control the launched power before the signal enters to the channel.

For the channel, two configurations were set up: optical back-to-back and inline-EDFA transmissions. In the optical back-to-back configuration, the VOA at the transmitter together with an EDFA before the coherent reception were used to vary SNR. With the inline EDFA-based fiber transmission, two distances were considered: 300 km and 500 km SSMF which consist of 3 and 5 in-line EDFAs (6 dB noise figure) - each after 100 km of SSMF (Sterlite G.652.D) for compensation of the fiber loss, respectively.

At the receiver, the optical signal was first converted into electrical domain using a homodyne coherent reception. It consists of a local oscillator (LO) (linewidth ∼100 kHz), a 90° hybrid and four pairs of balanced photo-detectors. Electrical signals were captured and digitized by a real-time oscilloscope with 8-bit ADCs sampling at 100 GSa/s before off-line processing. In the off-line DSP, re-sampling took place first at 2 samples per symbol. Then, the digital signals were formatted/scaled by a signal conditioning module. For the fiber transmission, the impact of chromatic dispersion (CD) was
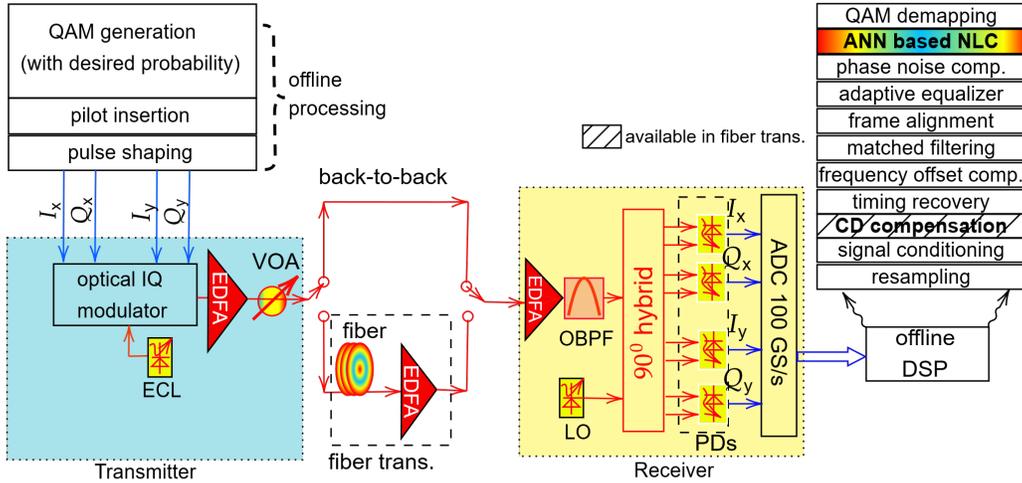
Fig. 4: Experimental setup for a dual polarization probabilistically-shaped 28 GBaud 64/256-QAM systems. ECL: external cavity laser, EDFA: Erbium-doped fiber amplifier, VOA: variable optical attenuator, OBPF: optical bandpass filter, LO: local oscillator, PDs: photodetectors, CD: chromatic dispersion, DSP: digital signal processing, ANN-based NLC: artificial-neural-network based nonlinearity compensation, trans.: transmission.

removed simply by using an inverse function of CD in frequency domain. The timing recovery and frequency offset error correction algorithms were placed before the matched-filtering using a Gardner phase detector and a conventional Fourier-transform-based method, respectively [23], [24]. A pilot-aided butterfly-structure adaptive equalizer (21 taps) was then carried out to cancel linear effects [25]. The well-known constant-modulus algorithm was used for adapting filter coefficients at pilot locations. Phase noise was estimated and compensated using a conventional pilot-aided (CPA) method. The CPA estimated the phase noise in a block-wise manner in which 8 pilots in each block was deployed for sufficient noise averaging [26]. After this stage, the ANN-based NLC was deployed to compensate for the nonlinear impairment from the transceivers and/or the optical fiber. Finally, just before QAM de-mapping, the pilot symbols were removed and only PCS 64/256-QAM symbols were taken into account for MI estimation with the aid of the transmitted data followed Eq. 2.

IV. RESULTS AND DISCUSSION

To assess the performance of the proposed systems, we adopt the figure of merit of MI given in bits/symbol. MI was measured for DP but their averages were presented. MI was computed from 15 frames, each of ∼60 000 symbols per polarization.

A. Optical back-to-back performance

Fig. 5 shows the system performance with and without the proposed ANN-based NLC in terms of MI versus equivalent SNRs (derived from optical SNRs in a reference bandwidth of 12.5 GHz) for DP 64-QAM (Fig. 5-a) and 256-QAM (Fig. 5-b) in the optical back-to-back configuration. For both modulation formats, a curve labelled as "Shannon" is also plotted for a reference.

For the shaped 64-QAM system (Fig. 5-a), there is only a little improvement (< 0.05 bits/symbol) for all systems under test, i.e. uniform, PCS with $\kappa = 0.03$ and PCS with $\kappa = 0.07$, when equipping the proposed scheme for transceivers nonlinearity compensation. This is due to the fact that the number of quantization bits of DAC/ADC in this experiment is large enough to support 64-QAM signals and the modulator was well-calibrated. This figure also shows that around more than 1 dB SNR shaping gain can be experimentally achieved using the two fixed PMFs in a wide range of SNRs, i.e $\kappa = 0.07$ for SNRs from ∼10 dB to ∼15 dB and $\kappa = 0.03$ for SNRs from ∼15 dB to ∼20 dB. Due to the little improvement when applying the NLC shown in 64-QAM, we focus on 256-QAM from now on for evaluating performance of the proposed NLC scheme.

On the other hand, the performance improvement when deploying 256-QAM with the ANN-based NLC is significant, as depicted in Fig. 5-b. It is evident that the SNR gains when deploying the ANN-based NLC are around 0.4 dB, 0.8 dB and 1 dB for PCS 256-QAM with $\kappa = 0$, $\kappa = 0.009$ and $\kappa = 0.019$ at the same MI of 5.5 bits/symbol, respectively. After deploying ANN-based NLC, around 2 dB SNR gain were experimentally achieved for PCS 256-QAM using the two aforementioned rates. This gain is larger than that of the 64-QAM systems because it comes from both shaping and NLC. The suitable SNR ranges associated with $\kappa = 0.019$ and $\kappa = 0.09$ for PCS 256-QAM are from ∼14 dB to ∼20 dB and from ∼20 dB to ∼26 dB, respectively.

At the same time, Fig. 5-b also provides information of the MI improvement when the ANN-based NLC is applied for the three studied 256-QAM systems on the right-hand y-axis, i.e. the different in MI between the filled-markers (with NLC) and opened-markers (without NLC) on the left-hand y-axis. Clearly, the MI enhancement under the same SNR is larger for deeper shaping systems, as can be seen from the MI
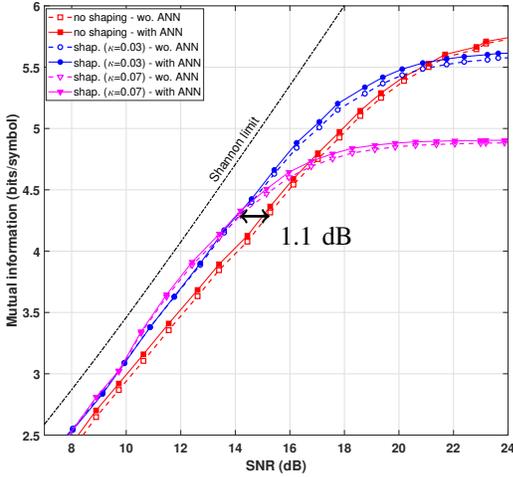
(a) 64-QAM
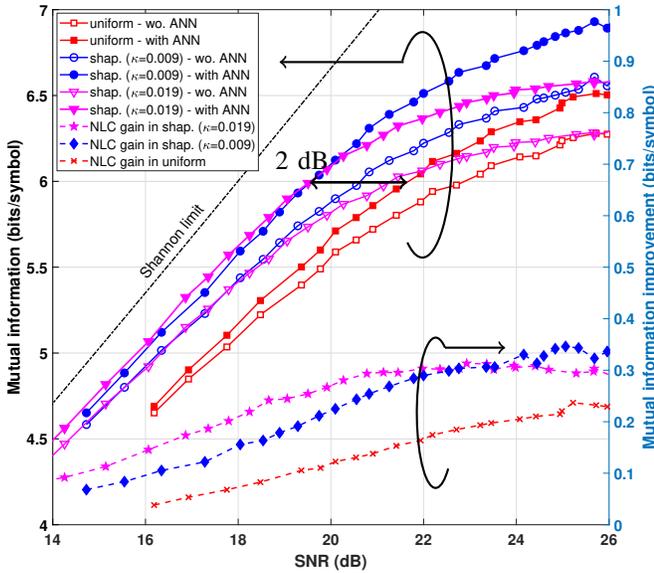
(b) 256-QAM
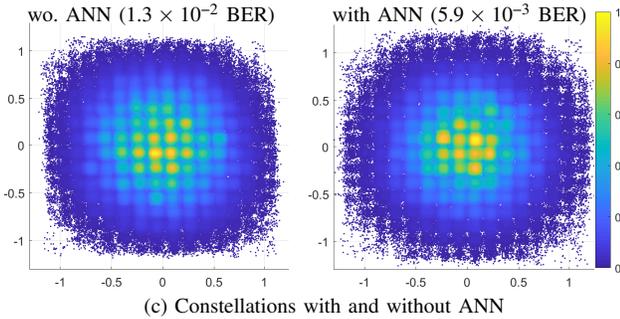
(c) Constellations with and without ANN

Fig. 5: Performance in terms of mutual information (MI) versus SNR of dual-polarization (DP) 28 GBaud transmissions with and without ANN-based NLC as a function of SNR under different shaping rates for (a) PCS 64-QAM and (b) PCS 256-QAM. (c) An example of constellation-diagram comparison between with and without ANN-based NLC for PCS 256-QAM, $\kappa = 0.019$ at 25 dB SNR. wo.: without.

improvement chart. The MI enhancement is larger for shaping systems in comparison to the uniform constellations. The maximum enhancement when deploying the ANN-based NLC for the uniform 256-QAM is $\sim$0.23 bits/symbol, whereas this value is above 0.3 bits/symbol for PCS 256-QAM, $\kappa = 0.019$. This chart clearly indicates that the more the shaping is, the larger improvement gains exist for systems equipped with the nonlinearity compensator, as shown in the star-marker curve ($\kappa = 0.009$) versus the diamond-marker curve ($\kappa = 0.019$) and cross-marker curve (no shaping, i.e. $\kappa = 0$) for up to 23 dB SNR. For high SNR regimes (> 24 dB SNR), however, this trend is not accurate for deeply shaped QAM. This is because the performance of deeper shaping system ($\kappa = 0.019$) is approaching its entropy for relatively high SNR regimes. The experimentally obtained results show a good agreement with the previous conclusion that deeper shaping imposes more nonlinear distortion from the transceivers. Fig. 5-c illustrates a comparison of constellation diagrams of PCS 256-QAM, $\kappa = 0.019$ at 25 dB SNR with and without using the nonlinear compensation. The ANN-based NLC compensates for transceivers nonlinear distortion which can be seen by comparing the constellation regimes with large amplitudes on this figure. In other words, decision boundaries especially at high power constellation points were corrected after ANN-based NLC. The accordingly measured bit-error-rate (BER) before and after the NLC are 1.3×10^{-2} and 5.9×10^{-3}, respectively.

B. Using fiber transmission

In this section, we experimentally evaluate the performance of the proposed NLC scheme with fiber transmission. Unlike to the optical back-to-back configuration, the ANN training in the fiber channel may take place multiple times at different launched powers or just once at the optimum launched power for each transmission distance. We thus investigate 3 different ANN configurations: (1) the trained ANN in optical back-to-back is also used for fiber transmissions, (2) the ANN is trained at the optimum launched power (4 dBm - Fig. 6) and then used for all launched powers, and (3) the ANN is re-trained at each launched power.

Fig. 6 shows performance of DP 256-QAM different ANN configurations after propagating through 300 km SSMF (subfigures a, b and c) and 500 km SSMF (sub-figures d, e and f). For each transmission distance, three systems are considered: $\kappa = 0$ (no shaping), $\kappa = 0.009$ and $\kappa = 0.019$. For each system (each sub-figure), performance of four scenarios are considered: without ANN-NLC and with ANN-NLC under three aforementioned training strategies.

For the uniformly distributed QAM, the results on Fig. 6-a and Fig. 6-d show clearly that there is no performance improvement when different training strategies are deployed for both transmission distances under test. This means that the nonlinear distortion comes mainly from transceivers but not from Kerr-induced fiber nonlinearity. Although there is a small improvement in nonlinear regime (high power) if the ANN is trained at different launched power in comparison with the ANN trained in the optical back-to-back, the current
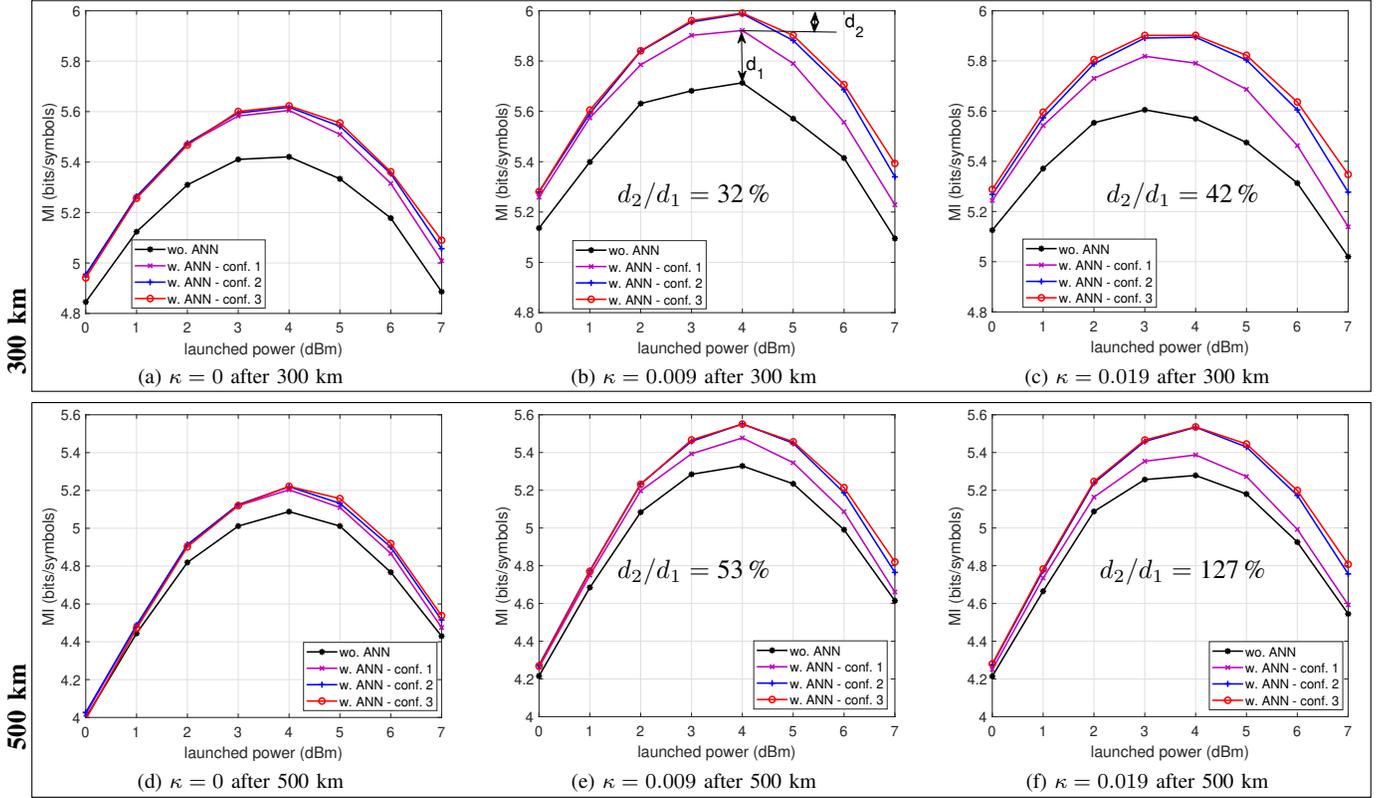
Fig. 6: Performance of the proposed ANN-based NLC for PS DP 256-QAM systems after 300 km and 500 km of fiber transmissions with different ANN configurations. (a) uniform constellation after 300 km, (b) PCS with $\kappa = 0.009$ after 300 km, (c) PCS with $\kappa = 0.019$ after 300 km, (d) uniform constellation after 500 km, (e) PCS with $\kappa = 0.009$ after 500 km and (f) PCS with $\kappa = 0.019$ after 500 km.
"wo. ANN": without ANN, "w. ANN - conf. 1": with ANN when the trained ANN in back-to-back is used for fiber transmission, "w. ANN - conf. 2": with ANN when the re-training is taken place once at 4 dBm and "w. ANN - conf. 3": with ANN when the re-training is taken place at every launched power.

ANN configuration is not sufficient to deal with pure fiber nonlinearity. In this uniform QAM transmission, the trained ANN from the back-to-back configuration thus can be reused in the fiber channel for dealing with transceiver nonlinearity.

In contrast, the trained ANN in the back-to-back does not give the optimum gain for the tested PCS systems in the fiber channel, as shown in Fig. 6-b,c,e,f. These figures show clearly that for the PCS systems, the ANN is either trained at the optimum launched power or trained at each launched power for an optimum gain and they also indicate that the proposed NLC scheme compensates at the same time transceivers and Kerr-induced nonlinearity for PCS systems. If the ANN is re-trained in fiber channel, at 4 dBm (optimum) launched power, MI is improved by ~0.3 bits/symbol and ~0.25 bits/symbol when deploying the proposed NLC scheme for PCS signals after propagating through 300 km (Fig. 6-b) and 500 km (Fig. 6-d) SSMF, respectively.

The results on Fig. 6-b,c,e,f indicates that the nonlinear functions/boundaries found via ANN-based NLC in the optical back-to-back is no longer close to the optimum ones under the impact of fiber nonlinearity for PCS systems. The additional gain, e.g. d_2 in Fig. 6-b, when the ANN is re-trained in the fiber channel depends on the shaping factor and the transmission distance. It can also be seen that if the ANN is re-trained in

the fiber channel, the additional gain increases according to the escalation of the shaping factor and the transmission link. This trend can be seen when comparing the ratio of the addition gain (d_2) provided if ANN is re-trained in the fiber channel to the gain provided if the ANN trained in the back-to-back (d_1). An example is shown in Fig. 6-b in which $d_2/d_1 = 32\%$. The ratio d_2/d_1 is smallest in Fig. 6-b and largest in Fig. 6-f. These ratios for Fig. 6-c,e,f are 42% 53% and 127%, respectively. These numbers of percentage also imply that more gain is achieved with the proposed NLC for (1) at same shaping factor but longer distance, i.e. more fiber-induced nonlinearity, and (2) at same distance but higher shaping factor, i.e. more nonlinearity due to higher PAPR. Thus, the coupled transceivers-fiber nonlinearity can be considered as a function of the shaping rate, transceivers' characteristics and the fiber's parameters. This interplay is complicated and it results in a considerable change of the nonlinear function/boundaries found in the optical back-to-back where only transceivers nonlinearity is taken into account. Therefore, the re-training is required to cope with that interplay. For all cases, the training at each power is not necessary to conduct as there is almost no improvement in comparison with the case of one-time training at the optimum launched power.

V. CONCLUSION

We have experimentally demonstrated, for the first time, the simultaneous ANN-based transceivers and Kerr-induced nonlinearity compensation for PCS DP 64/256-QAM optical systems. The effectiveness of the proposal was investigated for both optical back-to-back and fiber transmissions. A SNR gain of up to 1 dB improvement due to the transceiver NLC was obtained experimentally for shaping systems equipped with our proposed ANN-based NLC. On the other hand, a MI enhancement of $\sim$0.25 bits/symbol was achieved experimentally with our proposed scheme to combat with coupled transceivers-fiber nonlinearity in a fiber channel of up to 500 km SSMF for the tested PCS DP 256-QAM. Experimental results indicate that transceivers re-calibration may not be needed when a certain shaped-QAM is considered in the systems, e.g. 64 (or below)-QAM. However, the employment of additional DSP techniques such as the proposed NLC are necessary to compensate the coupled transceivers-fiber nonlinear distortion if PCS high-order QAMs is deployed, e.g. 256-QAM. For uniform 256-QAM, the training conducted during the calibration or in the optical back-to-back is not required to be repeated when fiber transmission is considered. However, such training is necessary for PCS signals to maximize the system performance, especially at a medium-to-long link and/or high shaping rates. For all cases, the re-training process with fiber transmission, if necessary, only takes place once at the optimum launched power when the proposed ANN-based NLC is used at the receiver.

REFERENCES

[1] M. P. Yankov, E. P. Da Silva, F. Da Ros, and D. Zibar, "Experimental analysis of pilot-based equalization for probabilistically shaped WDM systems with 256QAM/1024QAM," in *2017 Optical Fiber Communications Conference and Exhibition, OFC 2017 - Proceedings*. OSA, 2017, p. W2A.48.

[2] G. Bocherer, F. Steiner, and P. Schulte, "Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, 2015.

[3] J. Cho and P. J. Winzer, "Probabilistic Constellation Shaping for Optical Fiber Communications," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1590–1607, 2019.

[4] J. Cho, X. Chen, S. Chandrasekhar, and P. Winzer, "On line rates, information rates, and spectral efficiencies in probabilistically shaped QAM systems," *Opt. Express*, vol. 26, no. 8, pp. 9784–9791, Apr 2018.

[5] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *Journal of Lightwave Technology*, vol. 37, no. 2, pp. 230–244, Jan 2019.

[6] D. A. A. Mello, F. A. Barbosa, and J. D. Reis, "Interplay of probabilistic shaping and the blind phase search algorithm," *Journal of Lightwave Technology*, vol. 36, no. 22, p. 5096–5105, Nov 2018. [Online]. Available: http://dx.doi.org/10.1109/JLT.2018.2869245

[7] Z. Qu and I. B. Djordjevic, "Optimal constellation shaping in optical communication systems," in *2018 20th International Conference on Transparent Optical Networks (ICTON)*, July 2018, pp. 1–5.

[8] Q. Wang, Y. Yue, and J. Anderson, "Compensation of limited bandwidth and nonlinearity for coherent transponder," *Applied Sciences (Switzerland)*, vol. 9, no. 9, 2019.

[9] P. W. Berenguer, M. Nölle, L. Molle, T. Raman, A. Napoli, C. Schubert, and J. K. Fischer, "Nonlinear digital pre-distortion of transmitter components," *Journal of Lightwave Technology*, vol. 34, no. 8, pp. 1739–1745, April 2016.

[10] A.-R. Mahmood, S. Stylianos, P. Ian D., and F. Wladek, "Neural-network-based pre-distortion method to compensate for low resolution DAC nonlinearity," in *45th European Conference on Optical Communication, ECOC 2019 - Proceedings*. OSA, 2019, p. Th.1.B.4.

[11] T. T. Nguyen, T. Zhang, M. Abu-Romoh, and A. Ellis, "Artificial Neural Network-Based Compensation for Transceiver Nonlinearity in Probabilistic Shaping Systems," in *Optical Fiber Communications Conference and Exhibition (OFC) 2020*. OSA, 2020, p. W2A.44.

[12] T. Fehenberger, A. Alvarado, G. Bocherer, and N. Hanik, "On Probabilistic Shaping of Quadrature Amplitude Modulation for the Nonlinear Fiber Channel," *Journal of Lightwave Technology*, vol. 34, no. 21, pp. 5063–5073, 2016.

[13] J. Thrane, J. Wass, M. Piels, J. C. M. Diniz, R. Jones, and D. Zibar, "Machine Learning Techniques for Optical Performance Monitoring From Directly Detected PDM-QAM Signals," *Journal of Lightwave Technology*, vol. 35, no. 4, pp. 868–875, Feb 2017.

[14] S. M. Ranzini, F. Da Ros, and D. Zibar, "Joint low-complexity opto-electronic chromatic dispersion compensation for short-reach transmission," in *2019 IEEE Photonics Conference (IPC)*, Sep. 2019, pp. 1–2.

[15] E. Giacoumidis, A. Tsokanos, M. Ghanbarisabagh, S. Mhatli, and L. P. Barry, "Unsupervised support vector machines for nonlinear blind equalization in co-ofdm," *IEEE Photonics Technology Letters*, vol. 30, no. 12, pp. 1091–1094, June 2018.

[16] S. Zhang, F. Yaman, K. Nakamura, T. Inoue, V. Kamalov, L. Jovanovski, V. Vusirikala, E. Mateo, Y. Inada, and T. Wang, "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature Communications*, vol. 10, 2019.

[17] T. Nguyen, S. Mhatli, E. Giacoumidis, L. Van Compernolle, M. Wuilpart, and P. Mégret, "Fiber nonlinearity equalizer based on support vector classification for coherent optical ofdm," *IEEE Photonics Journal*, vol. 8, no. 2, pp. 1–9, April 2016.

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

[19] M. Schaedler, S. Calabro, F. Pittala, G. Bocherer, M. Kuschnerov, C. Bluemm, and S. Pachnicke, "Neural Network Assisted Geometric Shaping for 800Gbit/s and 1Tbit/s Optical Transmission," in *Optical Fiber Communication Conference*. Optical Society of America, 2020.

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

[21] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, Nov 1994.

[22] E. Giacoumidis, Y. Lin, J. Wei, I. Aldaya, A. Tsokanos, and L. P. Barry, "Harnessing machine learning for fiber-induced nonlinearity mitigation in long-haul coherent optical ofdm," *Future Internet*, vol. 11, no. 1, 2018.

[23] X. Zhou, "Efficient clock and carrier recovery algorithms for single-carrier coherent optical systems: A systematic review on challenges and recent progress," *IEEE Signal Processing Magazine*, vol. 31, no. 2, pp. 35–45, March 2014.

[24] M. Selmi, Y. Jaouen, and P. Ciblat, "Accurate digital frequency offset estimator for coherent polmux qam transmission systems," in *2009 35th European Conference on Optical Communication*, Sep. 2009, pp. 1–2.

[25] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Express*, vol. 16, no. 2, pp. 804–817, Jan 2008.

[26] S. T. Le, T. Kanesan, E. Giacoumidis, N. J. Doran, and A. D. Ellis, "Quasi-pilot aided phase noise estimation for coherent optical ofdm systems," *IEEE Photonics Technology Letters*, vol. 26, no. 5, pp. 504–507, March 2014.