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

Tu T. Nguyen, Tingting Zhang, Elias Giacoumidis, Abdallah A.I. Ali, Mingming Tan, Paul Harper, Liam P. Barry and Andrew D. Ellis

Abstract—In this paper, we experimentally demonstrate the combined benefit of artificial neural network-based nonlinearity compensation and probabilistic shaping for the first time. We demonstrate that the scheme not only compensates for transceiver's nonlinearity, enabling the full benefits of shaping to be achieved, but also the combined effects of transceiver and fiber propagation nonlinearities. The performance of the proposed artificial neural network is demonstrated at 28 Gbaud for both 64-QAM and 256-QAM probabilistically shaped systems and compared to that of uniformly distributed constellations. Our experimental results demonstrate: the expected performance gains for shaping alone; an additional SNR performance gain up to 1 dB in the linear region; an additional mutual information gain of 0.2 bits per channel use in the constellation-entropy limited region. In the presence of coupled transceiver and fiber-induced nonlinearities, an additional mutual information enhancement of $\sim$0.13 bits/symbol is experimentally observed for a fiber link of up to 500 km with the aid of the proposed artificial neural network.

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 [1]–[5]. Capacity demand means we need higher order QAM and every last dB of margin to maximally utilise the installed fiber plant. The principle behind PCS is to shape the signal constellation as close as possible to the optimum constellation for a given channel. For a Gaussian channel, a near-optimal signal-to-noise ratio (SNR) gain of 1.53 dB is feasible when employing PCS technique [2]. In optical communication, PCS is employed to limit the occurrence of high-power symbols in order to increase the SNR for all other symbols at the same mean power. At the optimum SNR, the slight loss in maximum capacity (entropy) from the reduced use of higher power symbols is more than offset by the SNR gain enjoyed by the lower power symbols. Although the PCS is one of the most promising candidates for next generation transponders, its impacts on the transceiver including fiber channel have not been explored in detail.

Generally speaking, the implementation of high-order modulation formats such as 64-QAM and beyond is often a big challenge due to the requirements for high SNR and high linearity of digital-to-analog converters (DACs)/analog-to-digital converters (ADCs). It is predicted that this problem may be more severe when PCS signals are modulated as a result of the enhancement of transceiver nonlinearity. This nonlinear enhancement comes from the fact that PCS signals generally have higher peak-to-average power ratio (PAPR) than that of uniformly distributed constellations [6]. In addition, PCS signals may require digital signal processing (DSP) adaptation for data recovery because conventional unsupervised-DSP algorithms are generally not compatible with shaping systems [1], [7].

The transceiver nonlinearity may be compensated using static digital filters (see [8], [9] for uniform QAM constellations), although it is difficult to estimate filter parameters due to mixing of nonlinear effects from different transceiver components such as DACs/ADCs and optical modulators. Under the impact of fiber nonlinearity, other nonlinear equalizers based on inverse Volterra series transfer functions may also be deployed to partially invert the nonlinear distortion induced by the transmission link. However, the Volterra-based nonlinear compensation (NLC) has shown worse performance than an optimized machine learning and their complexity is also high [10], [11]. Recently, a supervised machine-learning-based technique, namely artificial neural network (ANN), has been proposed and studied for uniform 64-QAM as a pre-distortion compensator for impairments induced by a low resolution DAC, but ignoring other nonlinear effects [12].

This same increase in PAPR also increases the susceptibility of PCS systems to fibre nonlinearity. Coupling of transceiver and fibre nonlinearity further complicates analytical descriptions due to the interplay of the various nonlinear terms, component bandwidths, and system memory (dispersion) of the multiple components of the link, which not only vary from link to link, but which may also vary with time. This

Tu T. Nguyen, Tingting Zhang, Abdallah A.I. Ali, Mingming Tan, Paul Harper and Andrew D. Ellis are with the Aston Institute of Photonic Technologies, Aston University, Aston St, Birmingham B4 7ET, UK. Emails (corresponding): t.nguyen14@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.

Liam P. Barry is with the School of Electronic Engineering, Dublin City University, Dublin 9, Ireland.

has drawn a great attention to applications of machine learning for fiber communications. Several common applications of machine learning in optical systems include performance monitoring and fiber-induced nonlinearity compensation [13]–[19]. Among them, the ANN is probably the most well-known algorithm that has been widely investigated in uniform QAM fiber communication systems. This is because the ANN is an effective solution for compensating both stochastic and deterministic distortion [15]. Moreover, it can be easily configured to operate as either nonlinear regression or classification. Machine learning-based algorithms, such as ANN and deep-learning algorithms, have also been implemented for the purpose of constellation design, also known as geometric shaping to improve tolerance to fiber impairments [20], [21]. The geometric shaping generally changes location of constellation points, whilst PCS assumes a fixed position for each constellation point but varies the probability of use. For practical implementation, conventional QAM grids such as 16-QAM and 32-QAM constellations are desirable due to their backward compatibility to matured DSP algorithms which have been developed for conventional optical systems. The geometric shaping is out of the scope of this work.

In this paper, we investigate for the first time the use of ANN to reduce implementation penalties due to coupled transceiver-fiber nonlinearity to enable the gains of high-order QAM and PCS to be more fully achieved. This work is extended from [22]. In [22], only transceiver's nonlinearity was taken into account while studying the effectiveness of an ANN-based NLC for PCS systems. The proposed ANN-based compensator is experimentally demonstrated for the PCS dual-polarization (DP) 28 GBaud 64/256-QAM system with different shaping factors (unshaped, moderate and high). Its performance is experimentally assessed for coupled transceiver nonlinearity, and coupled transceiver-fibre nonlinearity over 300 km and 500 km standard single mode fiber (SSMF) links employing inline Erbium-doped fiber amplifier with 100 km spacing. While there is no significant improvement observed in PCS 64-QAM, a SNR gain of 1 dB is experimentally demonstrated for mitigating transceiver nonlinearity for 256-QAM using the proposed ANN-based NLC scheme at a spectral efficiency (SE) threshold of 6 bits/symbol. For high regime of SNRs, the SE gain in terms of bits/symbol saturates at 0.2. For fiber transmission, by using the proposed NLC for compensating transceiver nonlinearity alone, the SE gain is retained for uniform 256-QAM with the distance of up to 500 km. However, considerable degradation of expected SE gain is observed for the shaping 256-QAM signals as the result of the enhancement of coupled transceiver-fiber nonlinearities. The ANN-based NLC thus requires to be re-trained for shaping systems to cope with this nonlinearity enhancement. An additional mutual information (MI) gain of ∼0.1 bits/symbol (total gain of 0.13 bits/symbol) is observed for the PCS 256-QAM system over a transmission distance of 500 km SSMF when the proposed ANN-based NLC is used to compensate for the coupled nonlinearity between the two transceivers and the fiber.

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

A. Probabilistic constellation shaping: a brief introduction

Constellation Shaping improves the performance of a linear transmission system by increasing the proportion of low energy symbols to reduce the mean power, whilst retaining a proportion of high energy symbols to minimise loss of capacity. In this work, this shaping is performed according to the the well known Maxwell-Boltzmann (MB) probability distribution (hence PCS), which is known to be the optimum for a linear additive white Gaussian noise channel, for generating a set of probability mass functions (PMFs). For uniformly distributed constellation points, the PMF $P_X(x_j)$ is given by [2]:

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

where x_j is one point of the input alphabet, M is the number of constellation points and κ is the shaping factor. For maximum transmission capacity, the shaping factor needs to be optimized as a function of signal power, modulation format and SNR [2]. We deployed square-QAM to simplify this optimisation to a one-dimensional (1D) optimization. Specifically we focus on 64-QAM and 256-QAM, corresponding to 8-pulse amplitude modulation (PAM) and 16-PAM, respectively. We also assumed that the location of shaped symbols are unchanged under the shaping, i.e. their real and imaginary amplitudes are integers of $\pm(2k+1)$, $k = 0, 1, ..., \sqrt{M}/2 - 1$, for calculating PMFs from Eq. 1 for PCS systems.

Fig. 1 shows the comparison in terms of MI in bit/symbol (representing the maximum amount of information bit in a symbol that can be transmitted with a vanishing probability of error, also known as symbol-wise achievable information rate [23]), between 64-QAM and 256-QAM under different
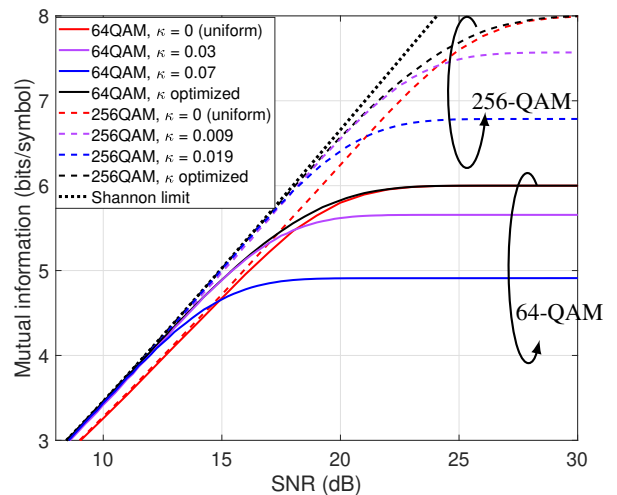

Fig. 1: Comparison of mutual information for 64-QAM (solid lines) and 256-QAM (dashed lines) with uniform ($\kappa = 0$, red lines), moderate shaping (purple lines), strong shaping (blue lines), and shaping optimised at each SNR for each format (black lines). Shaping factors for each format and strength shown in legend. Dashed line represents the Shannon limit.

shaping factors. Note that $\kappa = 0$ corresponds to the uniform constellation (no shaping). In this paper, MI is estimated by a Monte-Carlo simulation from N input-output symbol pairs (x_k, y_k). The input symbols are generated randomly according to the the PMF shown in Eq. 1. The output signals are the sum of the input signals and additive white Gaussian noise, such that the joint probability density function is $q_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{|y-x|^2}{2\sigma^2})$ with σ^2 being the noise variance. The MI is then calculated using the following equation [24]:

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

In Fig. 1, the Shannon capacity limit in Gaussian channel is also given as $\log_2(1+\text{SNR})$ for reference (labeled as "Shannon limit"). In addition to the unshaped case ($\kappa = 0$), four shaping rates are considered in this paper: $\kappa_1 = 0.07$ and $\kappa_2 = 0.03$ for PCS 64-QAM; and $\kappa_3 = 0.019$ and $\kappa_4 = 0.009$ for PCS 256-QAM. Fig. 1 shows numerically the MI evolution of 64/256-QAM under these 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 [25] in which only two fixed PMFs for each modulation format are sufficient for a wide 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 4.91, 5.66, 6.79 and 7.57, respectively. Fig. 1 also confirms numerically the maximum entropy and the SNR gain where the observed shaped performance is within 0.1 dB SNR of the maximum possible shaped performance. These results are summarised in Table I.

B. Impact of PCS signals on transceiver

It is clear that PCS changes the statistical properties of the transmitted signal. In this section, the PAPR metric is investigated for PCS signals with different shaping factors. The PAPR of a given signal $x(t)$ is the ratio of its peak power P_{peak} to its average power P_{avg} as PAPR (dB) $= 10\log_{10}\frac{P_{\text{peak}}}{P_{\text{avg}}}$.

Fig. 2 shows the statistics of PAPR in terms of complementary cumulative distribution function (CCDF) for different QAM signals under the three considered shaping factors. The vertical axis is the CCDF showing how often a certain PAPR value in the horizontal axis is exceeded. The CCDF curves on this figure were computed numerically from digital signal

TABLE I: Parameters of the PCS systems studied in this paper.

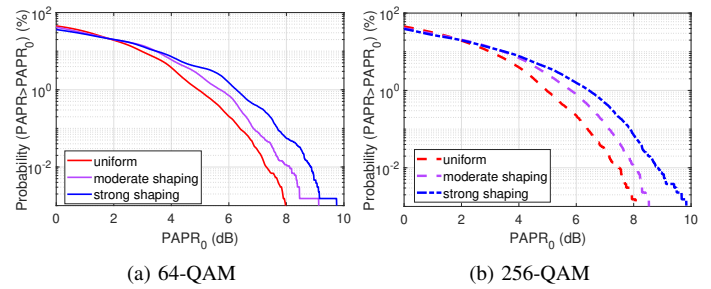| QAM	shaping type	κ	entropy	suitable SNR (dB)
64	strong	0.07	4.91	(-, 12]
64	moderate	0.03	5.66	[12, 16]
64	none	0	6	[21, -)
256	strong	0.019	6.79	(-, 18]
256	moderate	0.009	7.57	[18, 22]
256	none	0	8	[22, -)

Fig. 2: Numerically statistical quantity of PAPR in terms of complementary cumulative distribution function for (a) 64-QAM and (b) 256-QAM. Colors represent uniform (red), moderate shaping (purple) and strong shaping (blue).

after pulse-shape filtering (using a root-raised-cosine (RRC) filter with 0.1 roll-off factor and up-sampling at 2 samples per symbol). Simplicity, let $x(k)$, $k = 1, 2, ...N$ being the signal to measure probability of a given peak power, the CCDF of PAPR can be calculated through 4 steps: (1) computing the mean power $\mathbb{E}\{|x(k)|^2\}$ with $\mathbb{E}\{.\}$ being the expectation operator, (2) computing probability density function (PDF) of $10\log_{10}\frac{|x(k)|^2}{\mathbb{E}\{|x(k)|^2\}}$, (3) calculating cumulative distribution function (CDF) from the PDF, and (4) CCDF=1-CDF. Fig. 2 shows that the CCDF trends for both QAM formats are almost identical. The shaping signals exhibit larger PAPR than the uniform constellation cases ($\kappa = 0$) and the more shaping, the higher the PAPR. Specifically, at a probability level of 1%, the PAPR is 0.7 (1.5) dB higher with moderate (strong) shaping than without shaping for both 64-QAM and 256-QAM signals. The same amount of PAPR increment is also seen for 256-QAM signals. The PAPR increments result in an increase in the impact of transponder nonlinear distortions, unless the linear operation ranges are increased accordingly. The higher PAPR also imposes more Kerr-effect induced nonlinear noise which is proportional to the instantaneous power of the signal propagating through optical fiber.

C. Description of ANN-based NLC

Fig. 3 shows the structure of the simple feed-forward network deployed in this paper as the NLC, and is a typical ANN configuration. The aim of ANN algorithm is to find a function that maps the input to the desired target through a number of intermediate steps produced by neurons in the network. In general an ANN would comprise $4(m+1)$ real-valued inputs and 4 real-valued outputs with j hidden-layers of h_j neurons each. m represents the memory depth of the circuit, whereas the factor of 4 represents real and imaginary components of dual-polarization signals.

Within each neuron (Fig. 3-b), there are three calculations: (1) weight multiplication of inputs (subscripts $1, 2, ..., n$), (2) summing the weighted inputs and a bias (the input with subscript 0), and (3) passing the results of (2) through an activation function. In the proposed ANN-based NLC scheme, we selected a network with 5 memories ($m = 5$, i.e. current symbol with two symbols towards the pass and two symbols towards the future), two hidden layers ($j = 2$) and 10 neurons
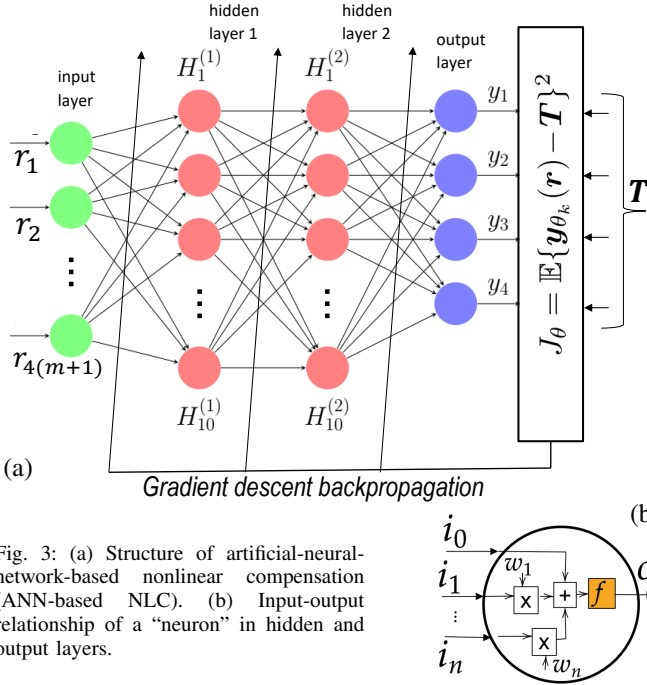
Fig. 3: (a) Structure of artificial-neural-network-based nonlinear compensation (ANN-based NLC). (b) Input-output relationship of a "neuron" in hidden and output layers.

per layer ($h_j = 10$ for all j). Thus, the input vector of the proposed NLC has 24 components, i.e. $4 \times (5 + 1) = 24$. 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 phase and the operational phase. As a supervised learning scheme, the training phase needs information of transmitted data. We used $\sim 28\,000$ and at least $300\,000$ QAM symbols in the training and the operational phase, respectively. For each configuration (format, shaping factor, fibre launch power), the training phase was carried out at initialisation. All the performance metrics were measured on the data in the operational phase using the same trained ANN.

The training phase is to optimize the weights and bias of the ANN. For this we used the batch Levenberg-Marquardt back-propagation algorithm (using gradient descent) [26]. $\sim 28\,000$ QAM symbols in the training phase were divided into two sets: $\sim 23\,000$ symbols (batch size) for the training set, which is used for computing the gradient and updating the weights and biases, and ~ 5000 for the validation set, which acts as pre-test to avoid over-fitting and under-fitting. At the very beginning of the training (state 0), initial weights, $\boldsymbol{\theta}_0$, are generated randomly. After the kth epoch (the number of epochs reflects the number of times that the learning algorithm will work through the entire training data set), application of the weights $\boldsymbol{\theta}_k$ to the received samples $\boldsymbol{r}$ gives an ANN output $\boldsymbol{y}_{\theta_k}(\boldsymbol{r})$, i.e. forward propagation. The difference between the ANN output $\boldsymbol{y}_{\theta_k}(\boldsymbol{r})$ and the expected symbols $\boldsymbol{T}$ (targets) is then used to calculate the loss function for that epoch $J_\theta = \mathbb{E}\{\boldsymbol{y}_{\theta_k}(\boldsymbol{r}) - \boldsymbol{T}\}^2$. Next, the gradients of the loss function (partial derivatives of the loss function with respect each weight) $\frac{\partial J_\theta}{\partial \theta_{pq}}$ are calculated in a back-propagation manner.
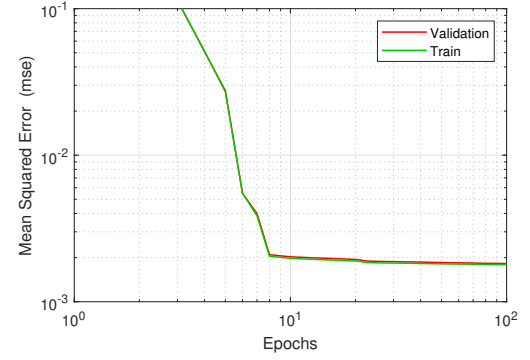

Fig. 4: ANN performance in terms of mean squared error versus number of epochs for training and validation sets for uniform 256-QAM.

The weight parameters are updated for the next epoch using $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \frac{\partial J_\theta}{\partial \theta}$. This cycle repeats on the same $23\,000$ symbols until the minima of the loss function (if validation performance increases more than 6 times since the last time it decrease) or the maximum number of epochs (100) is reached. Fig. 4 shows an example of the mean squared error (MSE) between the transmitted and received data as a function of number of epochs for uniform 256-QAM during the training and validation of the ANN. As clearly shown in this figure, the MSEs for both training set and validation set decrease when the number of epochs increases. The MSEs saturate at around $0.2\,\%$ after 20 epochs and that both MSEs are close to each other means no over/under-fitting issue with the training. After the training, the resultant optimum parameters and the fiber, are used with the ANN as the nonlinear compensator.

In the operational phase, the received signals are simply compensated by passing the received signals through the trained ANN without any further adaptation. It is reasonable to assume that the two nonlinearities are slowly varying in time (fiber nonlinearity due to amplifier output power drift and polarization mode dispersion evolution, and transmitter nonlinearity due to RF component aging). Thus, the training phase was only performed once at the optimum conditions, i.e. at the highest optical signal to noise ratio (OSNR) in back-to-back or at the optimum launched power in fiber transmissions. Whether or not the trained ANN in back-to-back can be deployed in the operational phase for fiber transmissions will be discussed more in detail in Section IV.

Regarding the complexity of the proposed scheme, this is a typical configuration of an ANN in which the detail analysis of the complexity of both training and operational phases could be found in [17]. In practice, the training process is typically taken place once and off-line. Due to that reason, the training process is normally not counted toward the implementation complexity for a practical system [16]. The number of real multiplications per symbol of the proposed ANN-based NLC operating in a forward manner is easily calculated as 380 real multiplications per symbol, i.e. $4 \times (5 + 1) \times 10 + 10 \times 10 + 10 \times 4 = 380$. Note that the activation function in each node can be implemented efficiently by using a look-up-table [15]. As the number of transmission configurations, i.e. the combination of modulation formats
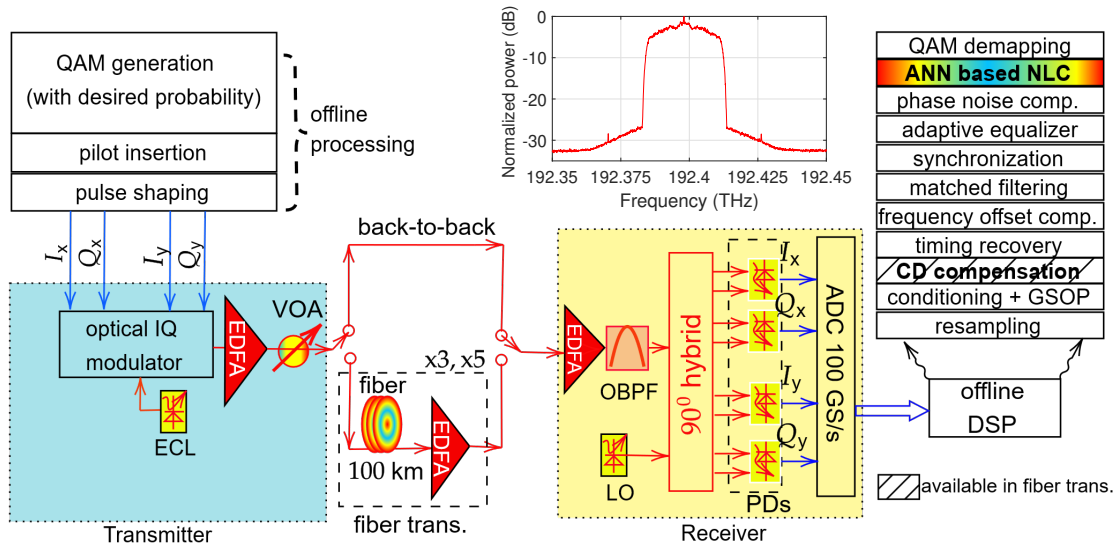
Fig. 5: Experimental setup for a dual polarization probabilistically-shaped 28 GBaud 64/256-QAM system. Inset: spectrum of uniform 64-QAM after the optical modulation. ECL: external cavity laser, EDFA: Erbium-doped fiber amplifier, VOA: variable optical attenuator, OBPF: optical bandpass filter, LO: local oscillator, PDs: photodetectors, GSOP: Gram–Schmidt orthogonalization procedure, CD: chromatic dispersion, DSP: digital signal processing, ANN-based NLC: artificial-neural-network based nonlinearity compensation, trans.: transmission.

(64/256-QAM), different channel conditions (back-to-back, 300 km and 500 km links) and different shaping factors, is large, the optimizations for each transmission configuration in terms of number of hidden layers, the number of neurons in each layer and the number of memories were not carried out in this work. We tested the ANN under some ANN's parameter sets for the uniform 256-QAM after 500 km fiber transmission and fixed at the nearly-optimum parameter set, i.e. 5 memories with 2 hidden layers and 10 nodes each layer, for all other configurations studied in this paper.

III. Experimental Setup

The experimental setup of the 28 GBaud PCS DP 64/256-QAM system is shown in Fig. 5. At the transmitter, four streams of 8/16-PAM data, each of $\sim 60\,000$ symbols, with desired PMFs according to the aforementioned shaping rates, were generated from four sequences of uniformly distributed pseudo-random bit sequences (part of the Gold sequence $2^{31} - 1$ with four different seeds: 5, 10, 15 and 20) using constant composition distribution matching [27]. 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 known pilot symbols (i.e. 1 pilot in every 20 symbols, equivalently 5 % pilot overhead) 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 because most of deployed DSP algorithms relied on pilot-aided symbols. The data was then up-sampled at 2 samples-per-symbol and pulse-shaped by a RRC filter with

a roll-off factor of 0.1. A special preamble of 9 ns (equivalently 256 symbols long), which consists of two repeated parts following [28], was inserted at the beginning of the payload to aid frame synchronization at the receiver. The signal was finally decomposed back to real and imaginary parts, scaled by its maximum amplitude and converted to integer form within the range of $[-127, 127]$ (8-bit DAC resolution). After this off-line processing, the signal was loaded into an Keysight arbitrary waveform generator (4-channel 8-bit DAC sampling at 56 GSa/s) and subsequently converted into the optical domain by applying the 600 mV pp signals to the inputs of a a commercial multi-format DP optical transmitter (laser linewidth ~ 100 kHz on 192.4 THz). An Erbium-doped fiber amplifier (EDFA) followed by a variable optical attenuator (VOA) was used to control the launched power. The inset of Fig. 5 shows an example of the spectrum of 28 GBaud DP 64-QAM signal after the optical modulator from which the drive amplitudes were estimated in the region of 55 % of Vpi.

Two configurations were set up: optical back-to-back and inline-EDFA transmission. For the optical back-to-back configuration, the VOA at the transmitter together with an EDFA before the coherent reception were used to vary OSNR. With the 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.

After the transmission, the optical signal was first converted into the electrical domain using an intradyne coherent receiver which includes 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 (8-bit ADCs) with sampling rate of 100 GSa/s before off-line processing, ten oscilloscope captures were collected for each OSNR in the optical back-to-back configuration

or for each transmission distance and each launch powers, giving frames of 60 000 symbols each oscilloscope capture.

The off-line DSP started with signal resampling at 2 samples per symbol. Then, the digital signals were formatted/scaled by a signal conditioning module. This module was also used to correct for the nonorthogonality between in-phase and quadrature components of the signal based on the Gram–Schmidt algorithm [29]. For the fiber transmission, the impact of static chromatic dispersion (CD) was removed simply by using an inverse function of CD in the 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 [30], [31]. After the matched-filtering, the frame synchronization was performed by using the Schmidl & Cox algorithm with the aid of the preamble [28]. For the channel equalization, a pilot-aided butterfly-structure adaptive equalizer (21 taps) was used. First, the filter coefficients were adapted at pilot locations using the well-known constant-modulus algorithm. Once the filter converged, it was applied for the whole signal to cancel linear effects and polarization demultiplexing [32]. 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 used for noise averaging [33]. After this stage, the ANN-based NLC was deployed to compensate for the nonlinear impairment from the transceiver 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 measurements followed Eq. 2. Effective SNR, which was measured from received constellations at the end of DSP chain, was also reported. Its reflects all transceiver impairment and the imperfection of the DSP chain. It was calculated from transmitted symbols and received symbols after the DSP chain as [34]

$$\text{SNR}_{\text{eff}} = \frac{\sum_{j=1}^{M} P_X(x_j)|\mu_j|^2}{\sum_{j=1}^{M} P_X(x_j)\sigma_j^2}, \tag{3}$$

where μ_j and σ_i are the mean and variance of the received symbols y_j that belong to the transmitted symbol x_j. For each oscilloscope, a frame length of $\sim$60 000 symbols per polarization was captured. Around the first half of this frame ($\sim$28 000 symbols) was used for training in the training phase whereas the second part of this frame was used in the operational phase to evaluate the performance of the proposed NLC scheme. The overhead used for training was negligible for the analysis because it took place one time at the highest OSNR condition in back-to-back or at the optimum launched power in the fiber transmission. Both MI and effective SNR were averaged from 9 oscilloscope captures, each of dual polarization.

IV. RESULTS AND DISCUSSION

A. Optical back-to-back performance

1) Implementation penalty: First of all, the implementation penalty was verified for both 64 and 256 PCS QAMs in the optical back-to-back setup when no NLC was used in the DSP

chain at the receiver. Shown in Fig. 6 is the effective SNR as a function of OSNR for all investigated PCS systems. The theory curve in this figure is derived from OSNR in a reference optical bandwidth of 12.5 GHz as $\frac{\text{OSNR} \times 12.5 \text{ GHz}}{\text{BW}}$ for dual polarization with BW being the signal bandwidth [35].

Fig. 6 indicates relatively same implementation penalty with <0.5 dB variation in the effective SNR for the two studied QAM formats suggesting that the impairments themselves are not dependent on the modulation format or shaping factor. For the low OSNR regime (<19 dB), the effective SNR increases linearly with the OSNR with the implement penalty of less than 2 dB. For the high regime of OSNR, however, this trend is no longer linear due to the imperfection of the DSP chain and the impact of the transceiver such as limited effective number-of-bits and limited bandwidth of DACs/ADCs, as well as nonlinear distortions. It is worth to mention that the implementation penalty of the uniform 256QAM is $\sim$0.5 dB more than that of the others shown in Fig. 6 due to the vulnerability of transceiver for high-order QAMs.

2) Mutual information gain obtained by ANN-based NLC: Fig. 7 shows the system performance with and without ANN-based NLC in terms of MI versus SNRs (derived from OSNRs) for DP 64-QAM (Fig. 7-a) and 256-QAM (Fig. 7-b). In this optical back-to-back experiment, for each QAM and at each shaping factor, the training phase was took place once at the highest OSNR condition. Then, the trained ANN was used for compensating the transceiver nonlinearity for all other OSNRs.

For 64-QAM without using the NLC (Fig. 7-a), around 1 dB SNR shaping gain is experimentally demonstrated for the SNR regime of <21 dB. As shown in the right-hand vertical axis of Fig. 7-a, which shows the difference in terms of MI between the system without NLC and the system with the aid of ANN-based NLC, little improvement (< 0.05 bits/symbol) is observed for all 64-QAM systems under test. However, the improvement is not the same for all studied shaping rates. As the SNR increases, the system performance becomes
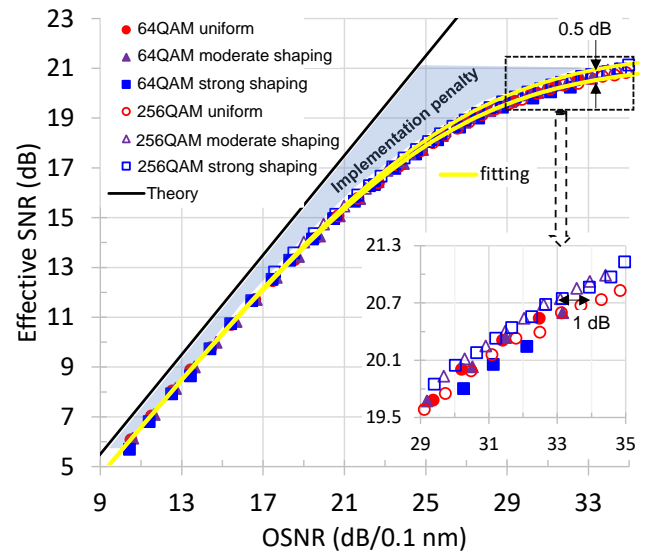

Fig. 6: Effective SNR versus OSNR (0.1 nm resolution bandwidth) for all studied constellation shaping 64/256-QAM systems.
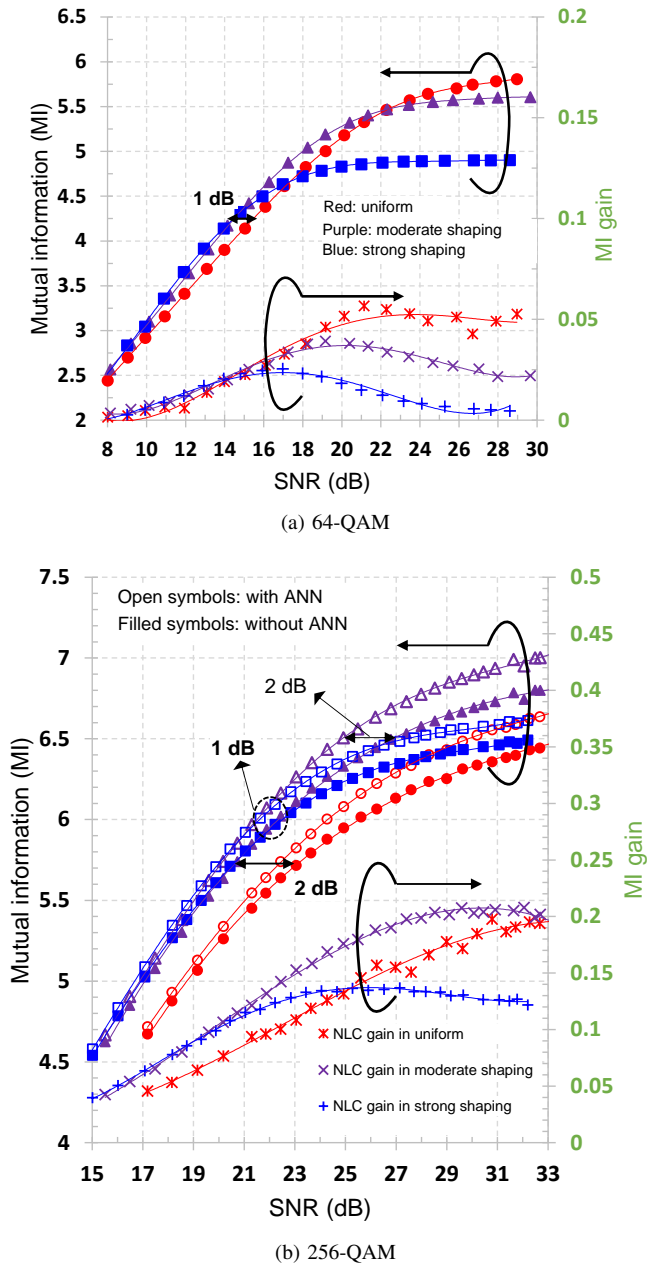
(a) 64-QAM

(b) 256-QAM

Fig. 7: Performance in terms of MI versus SNR (filled and/or open symbols, left-hand vertical axis) and MI gain provided by ANN (right-hand vertical axis) for probabilistic constellation shaping systems: (a) 64-QAM and (b) 256-QAM. Red, purple and blue symbols represent uniform, moderate and strong shaping, respectively. Open symbols and filled symbols are the MI with and without ANN, respectively. The MI-gain due to the ANN are presented in star, cross and plus markers for uniform, moderate and strong shaping, respectively. All lines represent curve-fitting.

increasingly dominated by the distortion from the transceiver, in this case, mainly the optical modulator and the DACs/DACs. The distortion is higher for constellation points far away from origin. Therefore, for weaker levels of shaping these higher amplitude points are more prevalent, and so the potential compensation gain is higher.

Fig. 7-b shows the performance of the 256-QAM under the three studied shaping factors as well as the MI gain obtained
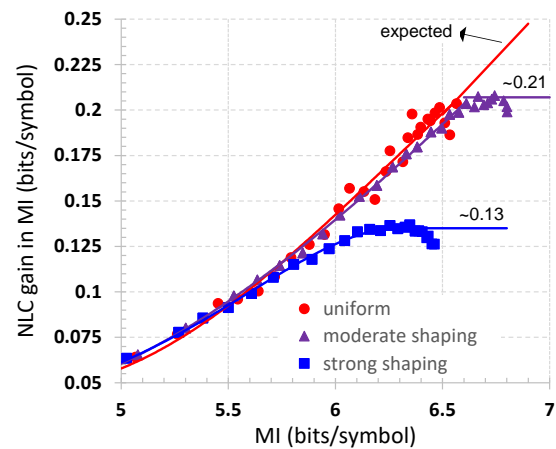

Fig. 8: ANN nonlinear compensation (NLC) gain in bits/symbol versus mutual information (MI) for 256-QAM uniform (red), moderate shaping (purple) and strong shaping (blue). All lines represent curve-fitting.

when the ANN-based NLC was used to compensate for the coupled nonlinearities at the receiver. Without using NLC, the shaped 256-QAM shows experimentally $\sim$2 dB SNR gain over the uniform 256-QAM at the same MI of 5.5 bits/symbol. Taking into account of $\sim$0.5 dB SNR extra implementation penalty for the uniform 256-QAM, the shaping gain alone is thus $\sim$1.5 dB SNR, which is higher than that of 64-QAM. This is what we expect for high-order PCS QAM (Fig. 1).

When deploying the proposed ANN-based NLC for 256-QAM, the performance improvement in terms of MI is significant – up to 0.2 bits/symbol as depicted in the right-hand vertical axis of Fig. 7-b. For SNRs below 25 dB, the gain provided by the proposed NLC for the shaped 256-QAM is clearly higher than that of the uniform one. The improvement for the strong and moderate shaping systems equipped with the NLC saturates at around 0.13 and 0.21 bits/symbol, respectively, whereas we do not see the gain saturation for the uniform 256-QAM, which is supposed to be higher than that of shaped ones as demonstrated in PCS 64-QAM. This is because we do not have high enough effective SNR for the uniform 256-QAM to reach its entropy. Fig. 8 plots NLC gain in terms of bits/symbol against MI for this 256-QAM under three studied shaping factors. Interestingly, at a same level of MI, the ANN-based NLC gives relatively the same SE improvement regardless of the tested shaping factors until the saturation point of each shaping factor. It suggests that one may easily estimate for a potential improvement when using the proposed NLC for any any shaping factors at a given SE. In terms of SNR gains, the proposed transceiver compensator gives $\sim$1 dB SNR gain for all studied 256-QAM at the same MI threshold of 6 bits/symbol. This SNR gain even goes as high as 2 dB at the threshold of 6.5 bits/symbol for the moderate shaping system, as illustrated in the left-hand vertical axis of Fig. 7-b. Whilst the trends in improvement due to the ANN are in line with expectations for both 64-QAM and 256-QAM, the level of improvement is low for the configuration employed in 64-QAM, and so for the fiber transmission, we focus on 256-QAM.

Fig. 9 visually shows an example of constellations with and
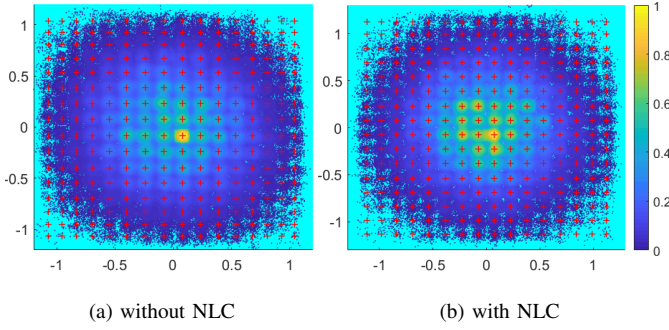
(a) without NLC   (b) with NLC

Fig. 9: An example of constellation-diagram comparison without (a) and with (b) ANN-based nonlinear compensation (NLC) for strong shaping 256-QAM at 32 dB SNR. Measured mutual information of (a) and (b) are 6.48 and 6.61 bits/symbol, respectively.
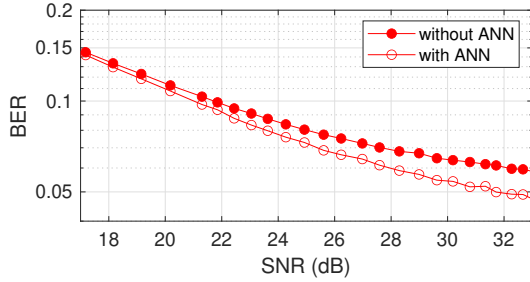

Fig. 10: Performance of uniform 256-QAM in terms of BER versus SNR with and without ANN.

without using the nonlinear compensator for strong shaping 256-QAM at the same 25 dB SNR. The NLC compensates for the transceiver nonlinear distortion which can be seen by comparing the constellation points with moderate-to-large amplitudes. In these constellations, the red crosses denote the distorted version of the transmitted constellation as the result of transceiver nonlinearity. They were computed as the mean values of received symbols that belong to each constellation point in the ideal constellation. As clearly shown in this figure, the distance on either real or imaginary part between two consecutive red crosses are no longer even, which is supposed to be for the transmitted constellation. For all experimental MI measurements in this work, the ideal constellation (i.e. transmitted alphabets) was first mapped to this distorted constellation before substituting into Eq. 2. It is worth to mention that there is no difference for effective SNR measurements in Eq. 3 using either distorted or ideal constellation. By comparing the received constellation with and without ANN (Fig. 9), it can be seen that the constellation with the aid of ANN is less "scattered", especially at the regimes with moderate-to-large amplitudes, leading to an improvement of ~0.13 bits/symbol. The Euclidean distances between the mean values of the constellations are more "even" for the case with ANN than that of the case without ANN. Another observation from Fig. 9 is that the constellation remains Gaussian-shaped distortion after NLC-ANN, which is not the case for some previous ANN's works [11], [17]. This may be explained by the observation that the nonlinear regression is smoother for higher order QAM targets, i.e. there are more steps existed in the targets for higher order QAM considered in this work.
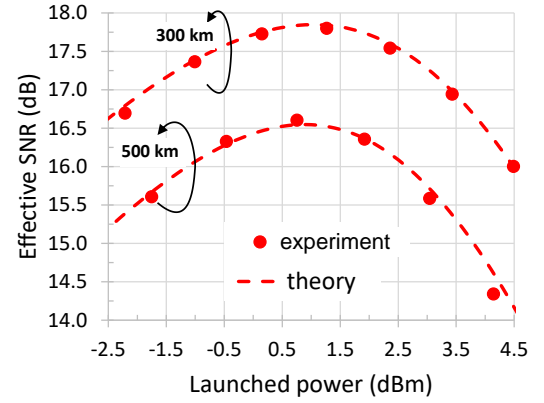

Fig. 11: Effective SNR versus launched power for uniform 256-QAM after 300 km and 500 km of fiber links.

Thus, the metric of MI and effective SNR used in this paper for performance assessment are reliable. To further validate this statement, the performance of uniform 256-QAM in terms of bit-error-rate (BER) versus SNR with and without NLC is shown in Fig. 10 as an example. It is easily seen that there is a good agreement of the SNR gain when the NLC is used for the uniform 256-QAM at the same BER threshold in Fig. 10 or at the same MI threshold in Fig. 7-b.

B. 256-QAM with fiber transmission

1) Transmission performance without NLC: For the fiber transmission, the performance of uniform 256-QAM without NLC was first verified for both tested transmission distances in terms of effective SNR versus launched power. The experimental results for these transmissions are shown in Fig. 11 in which experimentally theoretical expectation is also provided for comparison. It was calculated as:

$$\frac{1}{\text{SNR}_{\text{fiber}}} = \xi \frac{1}{\text{SNR}_{\text{GN}}} + \frac{1}{\text{SNR}_{\text{B2B}}}, \quad (4)$$

where SNR_{GN} is theoretical SNR for the fiber channel based on Gaussian-noise (GN) model [36]. $\text{SNR}_{\text{B2B}} = 21.5$ dB represents the maximum effective SNR in the optical back-to-back configuration. ξ coefficient accounts for the penalty implementation. $\xi = 1.8$ in these experiments, equivalently ~2.6 dB implementation penalty. Channel parameters used for SNR_{GN} calculation include; fiber's span length: 100 km, fiber's nonlinear coefficient: 1.3 W^{-1} km^{-1}, fiber's chromatic dispersion: 17 ps nm^{-1} km^{-1}, fiber's loss coefficient: 0.2 dB km^{-1}, gain of EDFA: 20 dB and noise figure of EDFA: 6 dB. The results in Fig. 11 shows a good match between theory and experiment in both linear and nonlinear regimes of the launched power. The maximum effective SNR is achieved at around 1 dBm for both tested distances.

In terms of achievable MI for all PCS 256-QAM with fiber transmissions, Fig. 12 shows the experimental comparison between the three studied shaping rates after 300 km and 500 km of fiber links. It can bee seen that the optimum launched power of ~1 dBm does not change for all tested PCS systems. For the same distance, the results in Fig. 12 shows
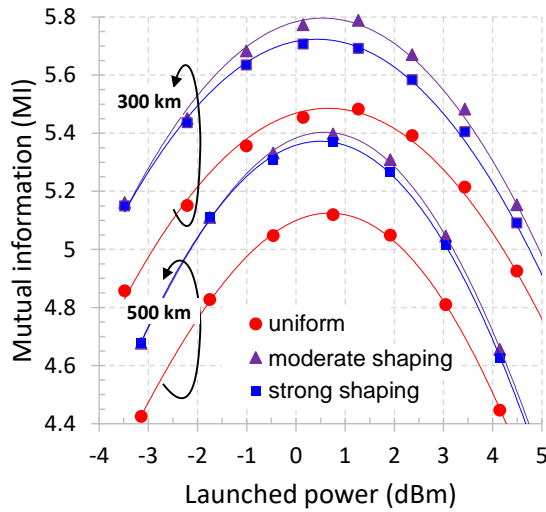
Fig. 12: Mutual information versus launched power for 256-QAM under different shaping factors after 300 km and 500 km of fiber links. Colors represent uniform (red), moderate shaping (purple) and strong shaping (blue). All lines represent curve-fitting.

clearly that the shaping systems always outperform the uniform one in terms of MI. While the moderate shaping shows a small MI advantage over the strong shaping over 300 km link, there is almost the same SE provided by them for the case of 500 km fiber transmission.

2) Performance with the aid of ANN-based NLC: The performance with the proposed NLC scheme was experimentally investigated for the two aforementioned distances. We considered two strategies for ANN training in the fiber transmission: (1) training performed back-to-back; and (2) training performed after transmission at the optimum launch power (∼1 dBm). For fiber transmission, the coupled transceiver-fiber nonlinearity may be dynamic. However, it is reasonable in practice to consider this dynamic behavior as a time slowly-varying (relatively stable over time) nonlinearity. Thus, we do not need to update the neurons in a dynamic way. So, the training here was also accomplished once for the fiber channel.

The performance difference (gain) in terms of MI for PCS 256-QAM when the proposed NLC was used with these training approaches is shown in Fig. 13 – left-hand charts and Fig. 13 – right-hand charts after propagating through 300 km and 500 km of SSMF links, respectively.

For uniformly distributed 256-QAM after 300 km, the results shown in Fig. 13 indicate clearly that there is almost no performance difference between the two training strategies at the optimum launched power of 4 dB. When the link increases to 500 km, a small NLC gain difference between two training strategies of ∼0.03 bits/symbol is observed for this uniform constellation.

In contrast, the big gaps between two different training strategies shown in Fig. 13 at the same power of ∼1 dBm for the moderate shaping and the strong shaping, imply that the trained ANN in the back-to-back does not give the optimum gain with fiber transmission for shaping systems. By re-training the ANN at the optimum launched power for each transmission distance, the NLC gains are further enhanced
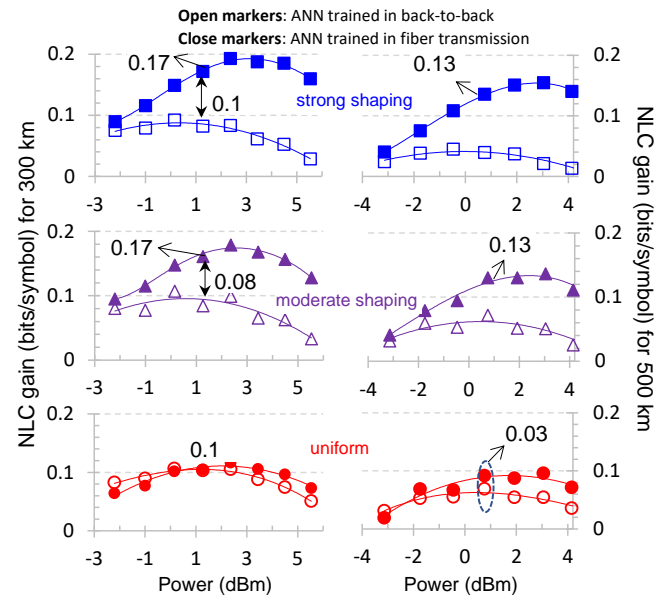

Fig. 13: Nonlinear compensation (NLC) gains in bits/symbol versus launched power for probabilistically shaped DP 256-QAM system after 300 km (left-hand charts) and 500 km (righ-hand charts) of fiber transmission. Red, purple and blue symbols represent uniform, moderate and strong shaping, respectively. Open symbols are the NLC gains when ANN trained in optical back-to-back, and closed symbols are the NLC gains when ANN trained at ∼1 dBm in fiber transmission. All lines are curve-fitting.

by ∼0.08 and ∼0.1 bits/symbol in the moderate and strong shaping transmission systems, respectively, for both studied distances. The reason could be coming from the considerable distortion as the result of the interaction between Kerr-effect fiber nonlinearities with shaping signals. The coupled transceiver-fiber nonlinearity can be considered as a function of the shaping rate, the transceiver's characteristics and the fiber's parameters. The nonlinear distortion is no longer dominated by the transceiver nonlinearity alone as in the uniform 256-QAM, but significantly changed in PCS systems. The stronger shaped signals, the larger PAPR becomes as discussed in section II, leading more Kerr-effect induced nonlinear noise. As a consequence, it requires the ANN to be re-trained in order to define a better nonlinear inverse function for fiber transmission with shaping signals. For both tested links, the additional gain provided ANN trained at the optimum launched power is higher for the stronger shaping QAM, indicating a strong impact of Kerr-effect induced fiber nonlinearities on shaping signals. This mean that the more the shaping is, the larger improvement gains exist for systems equipped with the proposed NLC compensator. By deploying the proposed NLC, a SE gain of ∼0.17 and ∼0.13 bits/symbol was achieved experimentally for the two tested shaping 256-QAMs over the link of 300 km and 500 km SSMF, respectively (Fig. 13).

In order to quantify how much gain coming from mitigating Kerr-effect fiber nonlinearity for all studied 256-QAM systems, the transceiver NLC gain obtained in the back-to-back (let it be Δ_{MI}^{trans}) is deducted from the coupled transceiver-fiber NLC gain (let it be Δ_{MI}^{coup}) for each training scheme. From results shown in Fig. 8, the expected transceiver NLC gain alone Δ_{MI}^{trans} for the three tested 256-QAM can be estimated at each MI
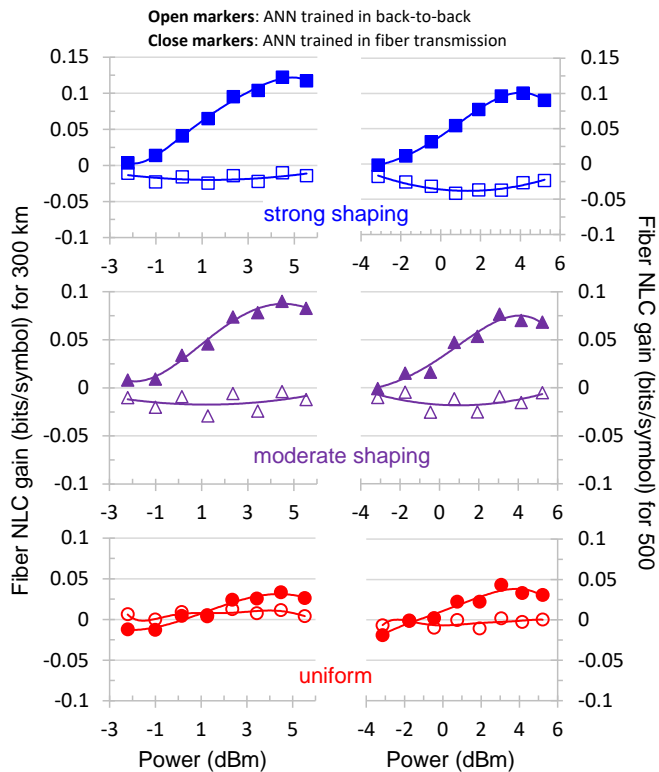
Fig. 14: Kerr-effect fiber nonlinear compensation (NLC) gain in bits/symbol for 256-QAM under different shaping factors after 300 km (left-hand charts) and 500 km (righ-hand charts) of fiber links. Red, purple and blue represent uniform, moderate shaping and strong shaping. Open symbols and filled symbols represent respectively the ANN trained in back-to-back and the ANN trained in fiber transmission at ∼1 dBm. All lines are curve-fitting.

level obtained for fiber transmission. The residual gain, i.e $\Delta_{MI}^{res} = \Delta_{MI}^{coup} - \Delta_{MI}^{trans}$, is used to evaluate the effectiveness of the ANN to deal with fiber nonlinearity. Fig. 14 shows these residual NLC gains against the launched power for 256-QAM, different shaping factors after 300 km (the left-hand charts in Fig. 14) and 500 km (the right-hand charts in Fig. 14). Starting with the residual gain provided by the ANN trained in the back-to-back (all open symbols), it can be seen that these open symbols are close to zero level for the uniform system but move gradually to below zero for the shaping ones. The most negative value is seen with the strong shaping system after 500 km, i.e. the largest shaping factor and the longest tested distance. In other words, when deploying the ANN trained in the back-to-back for the fiber transmission, the NLC transceiver gain is maintained for the uniform constellation. The pre-trained ANN in the back-to-back is thus can be used to compensate for transceiver nonlinearity in fiber transmission with uniform constellations. With the shaping signals, however, $\Delta_{MI}^{res} < 0$ indicates that the pre-trained ANN is not appropriate to be used for compensating transceiver nonlinearity in the present of Kerr-effect fiber nonlinearity. The transceiver nonlinearity that the ANN learnt in the back-to-back has been changed due to the interaction between the shaped signals and the fiber channel. The ANN thus needs to be re-trained in order to deal with this interaction.

When the ANN is re-trained in fiber transmission, Δ_{MI}^{res} represents NLC the gain coming from mitigating the Kerr-effect fiber nonlinearities. As shown in Fig. 14, generally, Δ_{MI}^{res} closes to zero in the SNR limited regime (low level of launched powers), and increases with the launched power for all tested transmissions. This is due to the reason that the coupled NLC gain is dominated by the transceiver nonlinearity at the linear regime. When the launched power increase, the Kerr-effect fiber nonlinearities becomes more important, and thus Δ_{MI}^{res} increases. The highest NLC rate is seen around ∼1 dBm where the ANN is trained at and the Kerr-effect fiber nonlinearities become more dominant. Above the optimum launched power, the Kerr-effect NLC gradually declines, and eventually reverses at ∼4 dB. It is worth to mention that in the uniform signal, the trained ANN at ∼1 dBm for fiber transmission shows worse performance than that of the ANN trained in the back-to-back in the SNR limited regime. This is due to the fact that the learnt transceiver-nonlinearity function is less accurate if the ANN trained after fiber transmission, compared to the case when ANN trained at highest OSNR condition in back-to-back. In any case, the trends are clear that more fiber nonlinearities due to the longer distances and/or stronger shaping signals, more benefit from training after transmission.

V. CONCLUSION

We have experimentally demonstrated, for the first time, the simultaneous ANN-based transceiver and Kerr-effect 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. The experimental results confirmed the expectation that deeper shaping imposes more nonlinear distortion from transceiver and Kerr-effect induced fiber nonlinearities, and thus leading to a higher NLC gain. For transceiver's nonlinearity alone, experimental results indicate that additional DSP may not be needed when a uniform QAM is replaced by a certain shaped-QAM under the same system's infrastructure, e.g. 64 (or below)-QAM with the devices involved in our study. However, the proposed NLC is worth to consider for compensating coupled transceiver-fiber nonlinear distortion when shaping signals and/or high-order QAMs are deployed, e.g. 256-QAM studied in this paper. For uniform 256-QAM, the ANN training conducted during the calibration or in the optical back-to-back was not required to be repeated when fiber transmission of up to 500 km was deployed. However, such re-training is necessary for shaping signals to cope with the complex interplay between fiber, transceiver and signal's properties, especially for medium-to-long links and/or strong shaping signals. Compared to the case of without using a NLC, the proposed NLC showed experimentally MI enhancement of ∼0.17 and ∼0.13 bits/symbol in a fiber link of 300 km and 500 km SSMF, respectively for the tested PCS DP 28 GBaud 256-QAM system. All of improvement was achieved without changing ANN configurations or shaping family, and that additional benefit may be possible with these modifications.

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] J. Renner, T. Fehenberger, M. P. Yankov, F. D. Ros, S. Forchhammer, G. Böcherer, and N. Hanik, "Experimental Comparison of Probabilistic Shaping Methods for Unrepeated Fiber Transmission," *J. Lightwave Technol.*, vol. 35, no. 22, pp. 4871–4879, Nov 2017.

[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] E. Giacoumidis, A. Matin, J. Wei, N. J. Doran, L. P. Barry, and X. Wang, "Blind nonlinearity equalization by machine-learning-based clustering for single- and multichannel coherent optical ofdm," *Journal of Lightwave Technology*, vol. 36, no. 3, pp. 721–727, 2018.

[11] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, and S. Pachnicke, "Deep neural network equalization for optical short reach communication," *Applied Sciences*, vol. 9, no. 21, p. 4675, Nov 2019.

[12] 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.

[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, 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.

[16] Q. Fan, G. Zhou, T. Gui, C. Lu, and A. P. T. Lau, "Advancing theoretical understanding and practical performance of signal processing for nonlinear optical communications through machine learning," *Nature Communications*, vol. 11, no. 1, p. 3694, Jul 2020.

[17] O. Sidelnikov, A. Redyuk, and S. Sygletos, "Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems," *Opt. Express*, vol. 26, no. 25, pp. 32 765–32 776, Dec 2018.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] A. Alvarado, T. Fehenberger, B. Chen, and F. M. J. Willems, "Achievable Information Rates for Fiber Optics: Applications and Computations," *J. Lightwave Technol.*, vol. 36, no. 2, pp. 424–439, Jan 2018.

[24] D. M. Arnold, H. . Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 8, pp. 3498–3508, Aug 2006.

[25] 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.

[26] 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.

[27] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 430–434, Jan 2016.

[28] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for ofdm," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, Dec 1997.

[29] I. Fatadin, S. J. Savory, and D. Ives, "Compensation of quadrature imbalance in an optical qpsk coherent receiver," *IEEE Photonics Technology Letters*, vol. 20, no. 20, pp. 1733–1735, Oct 2008.

[30] 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.

[31] 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.

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

[33] 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.

[34] R. Maher, A. Alvarado, D. Lavery, and P. Bayvel, "Increasing the information rates of optical communications via coded modulation: a study of transceiver performance," *Scientific Reports*, vol. 6, no. 1, p. 21278, Feb 2016.

[35] R. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, Feb 2010.

[36] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The gn-model of fiber non-linear propagation and its applications," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 694–721, 2014.