High-Capacity Coherent WDM Networks Empowered by Probabilistic Shaping and End-to-End Deep Learning

Ayam M. Abbass[1,2] and Raad Sami Fyath[1]

[1]*Al-Nahrain University, Baghdad, Iraq,*
[2]*Mustansiriyah University, Baghdad, Iraq*

https://doi.org/10.26636/jtit.2024.2.1482

Abstract — To optimize the functionality of coherent optical fiber communication (OFC) systems and enhance their capacity related to long-haul transmissions, wavelength-division multiplexing (WDM) and probabilistic constellation shaping (PCS) techniques have been used. This paper develops an end-to-end (E2E) deep learning (DL)-based PCS algorithm, i.e., autoencoder (AE) for a high-order modulation format WDM system that minimizes nonlinear effects while ensuring high capacity and considers system parameters, in particular those related to the OFC channel. Only the AE of the central channel is trained to meet the specified performance objective, as the system design employs identical AEs in each WDM channel. The simulation results show that the architecture should consist of two hidden layers, with thirty two nodes per hidden layer and a "32×modulation order" batch size to obtain optimal system performance, when designing AE using a dense layer neural network. The behavior of the AE is examined to determine the optimum launch-power ranges that enhance the system's performance. The developed AE-based PCS-WDM provides a 0.4 shaping gain and outperforms conventional solutions.

Keywords — autoencoder, coherent optical fiber communication, end-to-end deep learning, PCS-WDM system

1. Introduction

Due to the developments in coherent detection technologies, the capacity of standard single-mode fiber (SMF) based transmission systems has been increasing by 20% each year over the past few decades [1]. To address the escalating requirement for data rates in high-speed communication networks, wavelength-division multiplexing (WDM) has been implemented in nearly all communication networks [2]. The utilization of WDM can potentially enhance the capability of a single optical fiber to transmit several channels [3]. Because of the exponential increase in data traffic, high-order quadrature amplitude modulation (QAM) formats are acknowledged to optimize spectral efficiency within constrained bandwidths [4].

The effectiveness of transmission rates is limited mainly by nonlinear fiber optics and transceiver noise. This constraint becomes conspicuous when modulation formats of higher order are employed, which is generally associated with a greater rate of symbol error [5]. To optimize data rates in optical communication systems, an effective strategy is used to improve spectral efficiency by employing constellation shaping techniques. A shaping gain is achieved by probabilistic and geometric constellation shaping, compared to conventional QAM constellations [6], offering a noticeable gap in comparison with Shannon capacity [7].

Currently, numerous researchers are investigating PCS techniques, resulting in a significant reduction in the average constellation power [8]–[10]. This reduction has been achieved by enhancing the probability of inner constellation point occurrences while concurrently reducing the likelihood of occurrences of outer constellation points [11]. Regarding mutual information (MI), PCS can reduce the effects of fiber nonlinearity and improve system performance [12].

The utilization of machine learning (ML) in OFC systems has demonstrated its benefits in nonlinear compensation, performance monitoring, and modulation format recognition [13]. The use of autoencoders (AEs) and an end-to-end deep learning (E2EDL) strategy has shown that it is possible to optimize transceiver performance and simultaneously improve transmission efficiency [14]. AEs are unsupervised learning models that use input data as a supervisory signal. It forces the model output to match the input and the reconstructed output to match the input data [15].

Many studies have optimized E2EDL by considering the complex limitations of OFC networks. Studies have used feed-forward neural networks (FFNN) and sliding window bidirectional recurrent neural networks (SBRNN) for intensity modulation/direct detection (IM/DD) systems in AE-based networks. These efforts have shown potential to improve transmission performance [16]. Several researchers implemented a coherent OFC system based on E2EDL (i.e. AE), with the main objective of reducing the nonlinear fiber impacts without using the PCS technique [17]–[19].

Determining the most effective configuration for modulation formats and symbol probabilities in OFC remains a significant challenge and still is an unresolved issue [20]. Conventional approaches to constellation shaping assume that target distribution is symmetric to the origin for the additive white Gaussian noise (AWGN) channel. Nevertheless, symmetric probability distributions do not limit the application of

E2EDL in constellation shaping. They can rather be seamlessly integrated into a different channel model [12].

The authors in [21] addressed AE-shaped constellations probabilistically and provided an information-theoretic framework. This approach enabled learning symbol distributions and constellations that attained channel capacity with the AWGN channel. PCS for coded-modulation systems was introduced on the AWGN channel [22]. AE-based E2EDL is also used to optimize the PCS for multidimensional signals (PCS 4D-256-QAM) using the AWGN channel [23]. Further, an E2EDL-based PCS algorithm was implemented for a few-mode fiber (FMF) system utilizing the 5-WDM system, with 64-QAM across ten spans of 100 km each [12].

It is worth noting here that the reported E2EDL-based approaches have not been utilized for PCS-based high-capacity WDM systems operating in the presence of nonlinear fiber optics and using many channels with a high-order modulation format. PCS was investigated exclusively in single-channel optical communication, in an AWGN channel [21]–[23]. On the other hand, OFC used only a limited number of WDM channels for dual-polarization (DP) 64-QAM signaling [12]. Increased data rates achieved by long-haul coherent OFC systems require capacity enhancements in present and future optical communication networks.

This paper develops an E2EDL-based PCS algorithm for high-order modulation formats that minimize nonlinear effects while utilizing a high WDM system capacity at maximum transmission reach (MTR). The paper offers the following contributions:

1) Developing an E2EDL-based PCS for high-capacity WDM systems operating with 32, 64, and 96 channels, accessing the MTR and providing higher MI, using different symbol rates (20, 40, 60, and 80 Gbps).
2) Enhancing the performance of the E2EDL-based PCS-WDM by determining the best DL parameters, such as batch size and the number of hidden layers.
3) Assessing the developed AE by testing various dispersion values corresponding to each WDM channel.
4) Examining the behavior of the developed AE and identifying the best signal-to-noise ratio (SNR), MI, and bit error rate (BER) values that resulted in optimum launch power ranges for 32-, 64-, and 96- WDM systems.

The next sections are organized as follows. Section 2 presents the main concepts for the E2EDL-based PCS-WDM in a coherent OFC system. The simulation results and discussion are presented in Section 3. The parameters of the developed AE and its related DL are explained as well. Also, this section investigates the developed AE to identify the optimum launch power, study the effects of modulation formats and symbol rate on the performance of AE-based PCS-WDM, and to compare the performance of the AE-based PCS-WDM with that of a conventional system. Finally, conclusions derived from this study are given in Section 4.

This work has been implemented using the Python programming language and utilizing the TensorFlow framework to develop the simulation models.

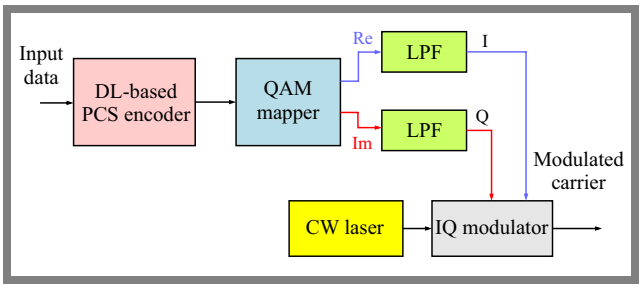

Fig. 1. Deep learning-based PCS channel transmitter.

2. E2EDL-based PCS-coherent Fiber WDM System

2.1. DL-based PCS Channel Transmitter

In a WDM system, each channel has its own transmitter and receiver operating at the channel wavelength and all the channels share the same transmission SMF link. A block diagram of the developed E2EDL-based PCS-channel transmitter is shown in Fig. 1. The transmitter comprises a DL-based PCS encoder, QAM-mapper, and an in-phase/quadrature (IQ) modulator with two low-pass filters (LPFs) inserted before it. A continuous-wave (CW) laser derives the modulator and acts as an unmodulated optical carrier. The binary encoded input data, produced by a pseudo-random number generator (PSRNG), is utilized as input for the encoder for the training process, with the distributed matcher (DM) generating the probability vectors which are fed to the QAM-mapper. Then, the output of the QAM-mapper consists of two distinct components, corresponding to the real and imaginary parts, each subjected to an LPF inserted to accomplish the pulse shaping operation. The outputs of the filters are applied to an optical IQ modulator, and the field of the channel CW laser is modulated, resulting in a modulated carrier.

A block diagram of the DL-based PCS channel encoder is shown in Fig. 2. The dense layers are used in an NN-based sampler NN_{ens} incorporating trainable weights (inclusive biases) w_{ens} that can be fine-tuned during training to enhance the system's performance. The encoder's role entails converting input data into an unnormalized log probability vector (i.e. logits) en with size (1 to M), where M represents the modulation order. The output of the encoder is fed to a softmax activation function to generate a discrete proba-

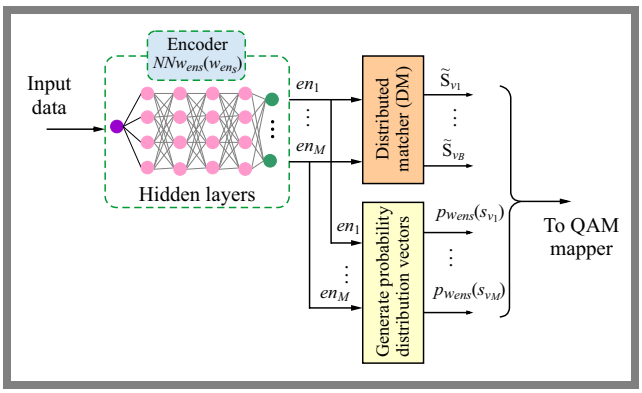

Fig. 2. DL-based PCS channel encoder.

bility distribution $p_{\boldsymbol{w}_{en_S}}(\boldsymbol{s_v})$ ranging from 1 to M, which is, in turn, applied to the QAM-mapper. At the transmitter, s_v is a one-hot vector that is modulated to constellation point (symbol) x by using the following function:

$$\boldsymbol{x} = f_{w_{en_M}}(\boldsymbol{s_v}) \ , \tag{1}$$

where $f_{\boldsymbol{w}_{en_M}}(.)$ is an NN_{en_S} based modulation with $\boldsymbol{w}_{\boldsymbol{en_S}}$ as trainable parameters. The NN_{en_S} consists of an input layer containing a single neuron, an output layer containing M neurons, and two hidden layers. The hidden layer consists of sixteen neurons, with each neuron using the rectified linear unit (ReLU) as an activation function. The primary goal of the DNN-based sampler is to identify the optimal probability distribution.

The process is executed comprehensively, whereby the channel condition plays a vital part. The channel condition is characterized by the nonlinear interference noise (NLIN) model which exhibits a dependency on the input power [24]–[26]. The NN-based sampler is trained to optimize the E2EDL loss function. The loss function incorporates the models of all the processing components, including the sampler, modulator, channel, and demodulator. The primary objective of the training procedure is to identify the optimal parameter configuration that effectively minimizes the loss function and improves the system's overall performance.

The output of the encoder $\boldsymbol{en_1}, \ldots, \boldsymbol{en_M}$ is fed into a distributed matcher (DM) that includes the Gumbel-softmax trick and a straight-through estimator [27]. This is used to overcome the limitation of the Gumbel-softmax trick that the resultant vectors $\tilde{\boldsymbol{s}}_{\boldsymbol{v_1}}, \ldots, \tilde{\boldsymbol{s}}_{\boldsymbol{v_B}}$ merely approximate the actual one-hot vectors $\boldsymbol{s}_{\boldsymbol{v_1}}, \ldots, \boldsymbol{s}_{\boldsymbol{v_B}}$, where B denotes the batch size that is used in the training process of AE-based PCS.

The complexity of training the proposed sampling mechanism for symbol s_v from a finite set S presents a challenge when utilizing ML-based PCS. This matter is resolved by applying the Gumbel-softmax trick [28], which is an expansion of the Gumbel-max trick [29]. As shown in [23] and [29], determining the maximizing argument of the sum of the sample with the Gumbel distribution g_i and $\log\left(p_{\boldsymbol{w}_{en_S}}(\boldsymbol{s_v})\right)$ is a practical approach to sampling a discrete distribution $p_{w_{en_S}}(s_v)$. The model used to calculate the samples is:

$$\boldsymbol{s_v} = \underset{i=1,\ldots,S}{\arg\max}\ g_i + \log\left(p_{\boldsymbol{w}_{en_S}}(i)\right) \ . \tag{2}$$

PCS is trained to utilize the optimization of an end-to-end loss function, which is predicated on the model encompassing all processing units, including the sampler. To minimize the loss function, the stochastic gradient descent (SGD) necessitates differentiable models for the modulator, detector, and channel [23]. Hence, developing a trainable sampler that exhibits differentiability regarding the parametric distribution $p_{\boldsymbol{w}_{en_S}}(\boldsymbol{s_v})$ presents a challenge when employing E2EDL-based algorithms for PCS. This entails excluding functions, such as $\max()$, $|.|$ and similar operations. Since the max operator lacks differentiability, the SGD method is inapplicable [12]. Consequently, the problem is resolved using the Gumbel-softmax trick which approximates $\max$ with $\mathrm{softmax}$ and generates an N-dimension vector, denoted by $\tilde{s}_v$. The mathematical expression for the Gumbel-softmax trick is given by [28]:

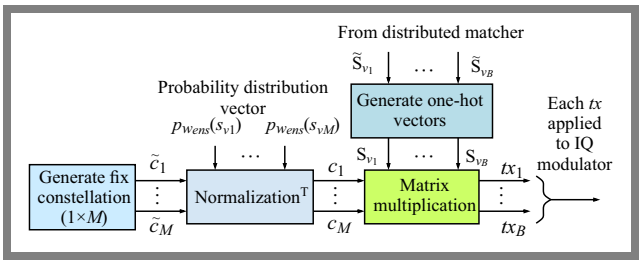

Fig. 3. Block diagram of a QAM mapper.

$$\tilde{\boldsymbol{s}}_{\boldsymbol{v}} = \frac{e^{\frac{g_i + \log\left(p_{\boldsymbol{w}_{\boldsymbol{en_S}}}(i)\right)}{\tau}}}{\sum_{j=1}^{S} e^{\frac{g_j + \log\left(p_{\boldsymbol{w}_{\boldsymbol{en_S}}}(j)\right)}{\tau}}}, i = 1, \ldots, N \ , \tag{3}$$

where τ is a positive parameter representing the softmax temperature. Note that samples from the Gumbel-softmax distribution become one-hot as τ approaches zero. By employing the straight-through estimator, the one-hot vector $\boldsymbol{s_v}$ is produced. This vector is then transformed into the QAM mapper. The transmitted symbol $\boldsymbol{tx}$, consisting of real and imaginary components, is subsequently subjected to separate processing via an LPF.

In conclusion, the filter outputs are applied to the optical IQ modulator, which effectively modulates the channel CW laser field and produces the modulated carrier. However, $\boldsymbol{s_v}$ is mapped into $\boldsymbol{tx}$ based on the using the formula:

$$\boldsymbol{tx} = f_{QAM}(\boldsymbol{s_v}) \ . \tag{4}$$

At the QAM mapper, a fixed constellation $\tilde{c}_1, \ldots, \tilde{c}_M$ is generated, as illustrated in Fig. 3. Subsequently, the constellation is normalized by employing the probability distribution vector $p_{\boldsymbol{w}_{en_S}}(\boldsymbol{s_v})$ and a transpose operation denoted as $c = [\tilde{c}_1, \ldots, \tilde{c}_M]^T$. Through the multiplication of a one-hot vector s_v by c, the transmitted constellation points are generated, each comprising real and imaginary components.

The modulation of these symbols utilizing an IQ modulator produces a carrier modulated before its transmission through the OFC channel. Normalization ensures that the anticipated energy of the constellation is exactly one.

2.2. Fiber Channel Model

The modulated carrier tx is transmitted via the OFC channel to yield the following output:

$$\boldsymbol{y} = f_{NLIN}(\boldsymbol{tx}) \ . \tag{5}$$

The NLIN channel model f_{NLIN} considers the impact of nonlinear interference on fiber communication [30], [31]. This model considers the launch power per channel and the moments of the constellation to capture the nonlinear effects that degrade the transmitted signal. The NLIN model simplifies these nonlinear effects as AWGN, with the variance

Tab. 1. Parameters of the simulated OFC-based PCS-WDM system.

System parameter	Value
Modulation format	DP 64-QAM
Number of WDM channels N_{ch}	32, 64, and 96
Symbol rate R_s	40 Gbps
Central channel frequency f_c	193.41 THz
Frequency channel spacing Δf	50 GHz
Number of link spans N_{sp}	MTR
Span length L	100 km
Fiber nonlinear coefficient Υ	1.3 (W·km)$^{-1}$
Fiber group velocity dispersion D at 1550 nm	16.5 ps/(nm·km)
Fiber dispersion slope $S \equiv \frac{dD}{d\lambda}$	0.08 ps/(nm^2·km)
Fiber attenuation α	0.2 dB/km
Optical amplifier gain G	20 dB
Optical amplifier noise figure	5 dB

determined by the parameters of the fiber communication channel. Consequently, the impairments in the channel are influenced by the amplified spontaneous emission (ASE) noise, which is governed by the amplifier noise figure F_n, the average launch power per channel and the high order moments of the constellation [32] are:

$$\mu_4 = \frac{\mathbb{E}[|X|^4]}{(\mathbb{E}[|X|^2])^2} \ , \ \ \mu_6 = \frac{\mathbb{E}[|X|^6]}{(\mathbb{E}[|X|^2])^3} \ . \tag{6}$$

The noise variance can be calculated as:

$$\sigma_n^2 = \sigma_{ASE}^2 \ F_n + \sigma_{NLIN}^2 \ , \tag{7}$$

where $\sigma_{ASE}^2 \ F_n$ is the ASE noise variance and σ_{NLIN}^2 is the nonlinear interference variance which is a function P_L, μ_4, and μ_6. Other parameters of the OFC-based PCS-WDM system are not included in Eq. (7) and are listed in Tab. 1.

2.3. DL-based Channel Receiver

The operation of the DL-based channel receiver is illustrated in Fig. 4. The received channel signal y is applied to the IQ demodulator, which utilizes a CW local laser to perform coherent demodulation. The demodulator generates an output $\tilde{y}$ comprising two components, namely real and imaginary. These two components are then subjected to a decoder for training purposes, with the ultimate goal of recovering the data that was initially transmitted. The recovered symbol de is obtained by passing $\tilde{y}$ through the decoder for training to recover the transmitted symbol according to:

$$de = f_{\boldsymbol{w}_{de_D}}(\tilde{\boldsymbol{y}}) \ , \tag{8}$$

where $f_{\boldsymbol{w}_{de_D}}(.)$ is a $NN_{w_{de_D}}$ based receiver with w_{de_D} as trainable parameters of the decoder.

Figure 5 shows the structure of the decoder located on the channel receiver side. The DNN-based sampler consists of

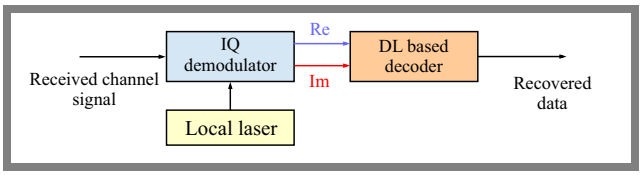

Fig. 4. DL-based PCS channel receiver.

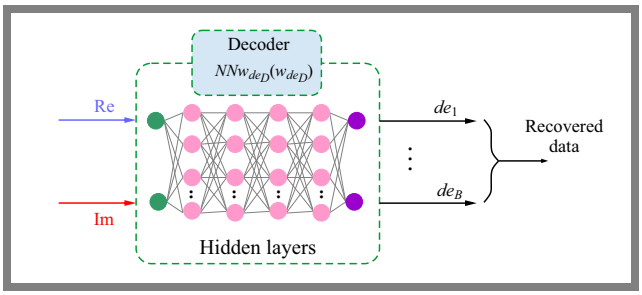

Fig. 5. DL-based decoder.

an input layer containing two neurons, an output layer containing m neurons, and two hidden layers, each with sixteen hidden neurons, and uses ReLU as an activation function. The output layer uses the softmax activation function. The probability vector generated by the softmax function represents the anticipated probability of each element from the input message that has been transmitted.

The mapping should be learned by the decoder, expressed as $\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{s_v}|\boldsymbol{y})$, which is approximately the true posterior distribution $p_{w_{en_S}}(\boldsymbol{s_v}|\boldsymbol{y})$ [12]:

$$\begin{aligned}
&L(w) = E_{\boldsymbol{s_v},\boldsymbol{y}}\left(-\log(\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{s_v}|\boldsymbol{y}))\right) \\
&= -\sum_{s_v=1}^{M} p_{\boldsymbol{w}_{en_S}}(\boldsymbol{s_v}) \int_y p(\boldsymbol{y}|f_{QAM}(\boldsymbol{s_v})) \log\left(\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{s_v}|\boldsymbol{y})\right) dy \\
&= -\int_{tx} p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx}) \int_y p(\boldsymbol{y}|\boldsymbol{tx}) \log\left(\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{tx}|\boldsymbol{y})\right) dydx \\
&= \int_{tx}\int_y p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx},\boldsymbol{y}) \log\left(\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{tx}|\boldsymbol{y})\right) dydx \\
&= -\int_{tx} p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx}) \log\left(p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx})\right) dx \\
&- \int_{tx}\int_y p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx}|\boldsymbol{y}) \times \log \frac{\tilde{p}_w(tx,y)}{p_{\boldsymbol{w}_{en_S}}(y) p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx})} dydx \ ,
\end{aligned} \tag{9}$$

where $\boldsymbol{w} = \boldsymbol{w}_{\boldsymbol{en_S}} + \boldsymbol{w}_{\boldsymbol{de_D}}$.

Equation (9) applies according to:
$p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx},\boldsymbol{y}) = p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx})p(\boldsymbol{y}|\boldsymbol{tx})$ and to
$p_{\boldsymbol{w}_{en_S}}(\boldsymbol{y}) = \int_{tx} p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx}|\boldsymbol{y})$ and
$\tilde{p}_w(\boldsymbol{tx},\boldsymbol{y}) = \tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{tx}|\boldsymbol{y})p_{\boldsymbol{w}_{en_S}}(\boldsymbol{y})$.

Here, $p_{\boldsymbol{w}_{en_S}}(\boldsymbol{tx},\boldsymbol{y})$ is a distribution of a true joint of (tx,y) and $\tilde{p}_w(\boldsymbol{tx},\boldsymbol{y})$ considers joint distribution in accordance with $\tilde{p}_{\boldsymbol{w}_{de_D}}(\boldsymbol{tx}|\boldsymbol{y})$. To get the optimum $\boldsymbol{w}$, iteratively minimizing the loss function as stated in Eq. (9) based on the stochastic gradient descent (SGD) method according to:

$$\boldsymbol{w}^{(j+1)} = \boldsymbol{w}^j - \eta \nabla_w \tilde{L}\left(\boldsymbol{w}^j\right) \ , \tag{10}$$

where η denotes the learning rate, j is the iteration number and $\nabla_{\boldsymbol{w}}\tilde{L}(\boldsymbol{w}^j)$ is an estimated gradient of the loss function. Accordingly, Eq. (9) can be stated as:

$$L(\boldsymbol{w}) = H_{w_{enS}}(\boldsymbol{tx}) - I_{w_{enS}}(\boldsymbol{tx}, \boldsymbol{y}) + E_y\left[D_{KL}(p_{\boldsymbol{w}_{enS}}(\boldsymbol{tx}|\boldsymbol{y})|| \times \tilde{p}_{\boldsymbol{w}_{deD}}(\boldsymbol{tx}|\boldsymbol{y}))\right] , \quad (11)$$

where $H_{w_{enS}}(\boldsymbol{tx})$ is the entropy of $\boldsymbol{s_v} = (s_{v_1}, \ldots, s_{v_B})$. MI between the input of the OFC channel $(tx_1, \ldots, tx_B)$ and its output $(y_1, \ldots, y_B)$ is denoted by $I_{w_{enS}}(\boldsymbol{tx}, \boldsymbol{y})$, $D_{KL}\left[p_{w_{enS}}(\boldsymbol{tx}|\boldsymbol{y})|| \times \tilde{p}_{\boldsymbol{w}_{deD}}(\boldsymbol{tx}|\boldsymbol{y})\right]$ is the Koulback-Liber (KL) divergence between true $p_{\boldsymbol{w}_{enS}}(\boldsymbol{tx}|\boldsymbol{y})$ and approximated $\tilde{p}_{\boldsymbol{w}_{deD}}(\boldsymbol{tx}|\boldsymbol{y})$ posterior distributions.

Optimizing MI and minimizing KL divergence are related to minimizing the loss function. Reducing divergence by approximating the true posterior distribution, the DNN-based detector approximates the maximizer of MI.

Each channel receiver uses a DP coherent detection scheme to extract the data from the received modulated optical carrier corresponding to that channel. The receiver operates with a 45-degree polarized CW local laser, with the frequency matching that of the unmodulated carrier at the channel transmitter. The local laser output goes through a polarization splitter to yield two equal-power orthogonal components acting as the local fields for the two polarization versions of the constellation shaping-channel receivers.

Further, each channel receiver has its own DSP to estimate the BER and MI of that channel. This DSP does not go through complicated computations to compensate for the linear and nonlinear effects of the fiber channel, since this job is already done during the ANN-training operation.

The SNR is a metric that incorporates various sources of noise, including amplification noise, nonlinear effects, and other defects in the transmitter. In this work, an ideal transmitter is assumed. The expression for SNR is:

$$SNR = \frac{\sigma_s^2}{\sigma_n^2} = \frac{\sigma_s^2}{\sigma_{ASE}^2 + \sigma_{NLI}^2} , \quad (12)$$

where σ_s^2 and σ_n^2 denote the transmitted (i.e. launch) and the total noise power, respectively. σ_{ASE}^2 is the variance of the noise generated by the amplification stages of the EDFA and σ_{NLI}^2 represents the noise variance caused by nonlinear interference (NLI), which includes both the intra- and the inter-channel distortions.

BER is a performance metric that quantifies the likelihood of an error by the number of incorrect bits per transmitted bit [34]. The BER of M order modulation format is calculated by [35]:

$$BER = \frac{2}{m}\left(1 - \frac{1}{\sqrt{M}}\text{erfc}\sqrt{\frac{3m(SNR)}{2(M-1)}}\right) , \quad (13)$$

where M is the number of discrete symbols involved in the modulation (i.e. modulation order), m is the number of bits per transmitted symbol ($m = \log_2 M$), and erfc denotes the complementary error function.

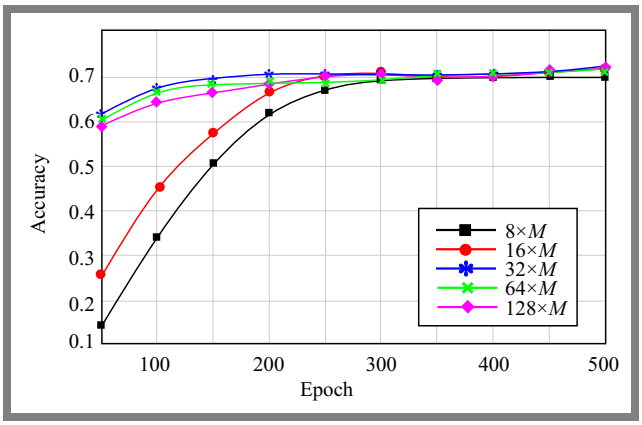

Fig. 6. Variation of accuracy with epoch numbers for DP 64-QAM, $N_{ch} = 32$ with $N_{sp} = 22$, and $P_L = -1$ dBm.

3. Simulation Results and Discussion

An analysis of transmission-related performance is conducted to determine the MTR, SNR, MI, and BER for various DP-QAM signaling and R_s values. The optical link comprises 100 km SMF spans. To mitigate OFC loss, an optical amplifier is inserted after each span. ITU wavelength grid standard dictates that when $R_s = 20, 40, 60$, and 80 Gbps, Δf can be set to 25, 50, 75, and 100 GHz, respectively. A threshold BER of $3.8{\times}10^{-3}$ is considered, corresponding to 7% hard decision FEC coding.

In this study, the AE is trained using predetermined values for modulation format, channel launch power P_L, and transmission distance. The effectiveness of the learned constellations is then evaluated utilizing the NLIN model. The simulation is executed within TensorFlow to train the AE. The simulated AE utilizes the WDM system parameters detailed in Tab. 1.

3.1. Effect of DL Parameters on AE Performance

With batch size serving as an independent parameter, Fig. 6 illustrates the variation of the accuracy of the developed AE. For DP 64-QAM signaling ($M = 64$), the outcomes are

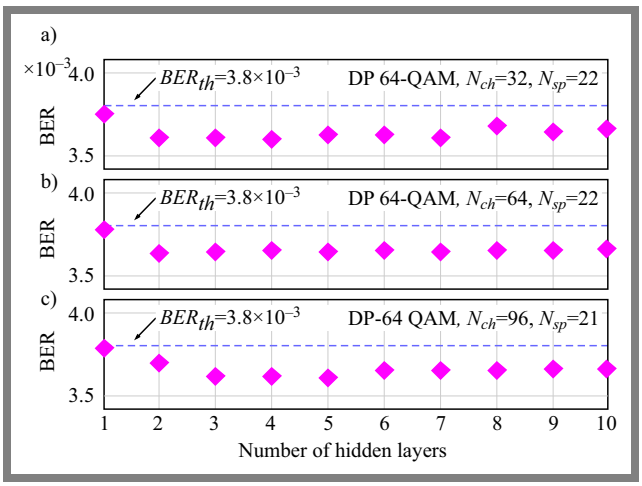

Fig. 7. Variation of BER with several hidden layers for DP 64-QAM. WDM system operating with $N_{sp} = 21$, $P_L = -1$ dBm, and $R_s = 40$ Gbps: a) $N_{ch} = 32$, b) $N_{ch} = 64$, and c) $N_{ch} = 96$.

acquired for a WDM system functioning with the number of WDM channels $N_{ch} = 32$, symbol rate $R_s = 40$ Gbps, launch power $P_L = -1$ dBm, number of link spans $N_{sp} = 22$ for DP 64-QAM signaling. Figure 6 shows an integer value x of modulation order M for batch size B. The AE training procedure incorporates a range of epochs and diverse batch sizes to identify the batch size that optimizes accuracy, thereby minimizing losses. When training the AE with batch sizes of 32×M and 64×M, as illustrated in Fig. 6, the accuracy attains its peak value at epoch 200.

The best number of hidden layers N_{hl} that results in the lowest BER for DP 64-QAM is determined based on the results provided in Fig. 7. The considered scenarios are N_{ch} =32, 64, 96, with $N_{sp} = 21$ and $P_L = -1$ dBm. The primary conclusions drawn from the data suggest that utilizing a structure with two hidden layers and a learning rate of 0.01 in the AE platform results in a minimum BER of 3.61×10^{-3} and 3.64×10^{-3} for N_{ch} =32 and 64, respectively. Additionally, a structure with four hidden layers yields a BER of 3.61×10^{-3} for $N_{ch} = 96$. All these BERs are below 3.8×10^{-3} which corresponds to the threshold value BER_{th}, associated with the 7%-HD FEC. Therefore, this study confirms the appropriateness of the chosen N_{hl} for the system under investigation.

It is worth mentioning that there is practically not much difference in performance depending on the number of hidden layers applied. In a low-complexity design, one may go with 1-2 layers, as suggested by the reviewer.

Note that the training of AE is implemented on the central channel of the WDM system in this work. The acquired AE parameters are subsequently applied to the AEs of other channels. This approach simplifies and accelerates the training process for AEs and, at the same time, guarantees that all channels satisfy BER performance prerequisites.

Alternatively, in a time-consuming approach, one may go to train the AE at each channel index for DP 64-QAM, $N_{ch} =$ 32, 64, and 96, with $N_{sp} = 22$, 22, and 21, respectively, and $P_L = -1$ dBm. The fiber group velocity dispersion value D is then computed for each channel using:

$$D(\lambda_i) \equiv D_i = D(\lambda_{ref}) + (\lambda_i - \lambda_{ref})S\,, \qquad (14)$$

where λ_i is the wavelength of the required channel, $\lambda_{ref} =$ 1549.71 nm is the wavelength of the reference (central) channel, where $D = 16.5$ ps × (nm·km)$^{-1}$, and $s \equiv dD/d\lambda$ represents the dispersion slope = 0.08 ps × (nm^2 km)$^{-1}$.

AE training using the two approaches renders almost the same results regarding SNR, MI, and BER. Figures 8 and 9 show the variation of BER and (SNR, MI) with channel index, respectively, using the same parameters from Fig. 7. The gain of the EDFA is assumed to be constant among C-band WDM channels, which is a reasonable assumption for the C-band system. Going to a multi-band WDM system incorporating more bands beside the C-band, more realistic scenarios could include the EDFA gain spectrum. These results demonstrate the consistency and dependability of the designed AE in both approaches.

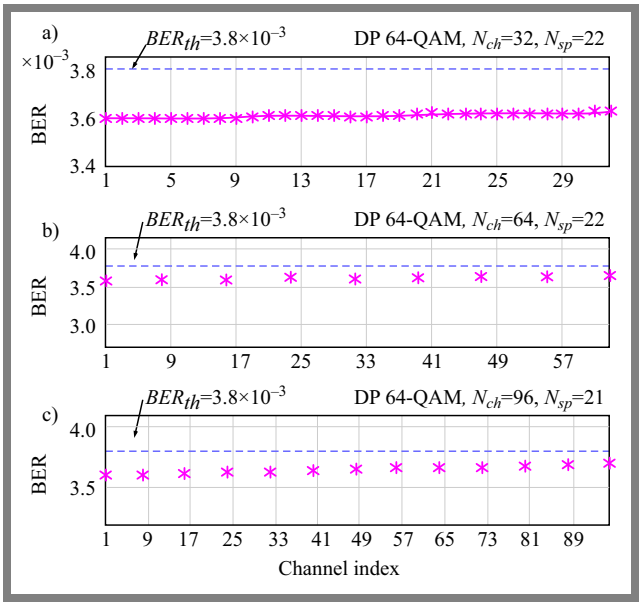

Fig. 8. BER on channel index for DP 64-QAM in WDM system operating at $P_L = -1$ dBm and $R_s = 40$ Gbps: a) $N_{ch} = 32$, b) $N_{ch} = 64$, and c) $N_{ch} = 96$ with $N_{sp} = 22, 22, 21$, respectively.

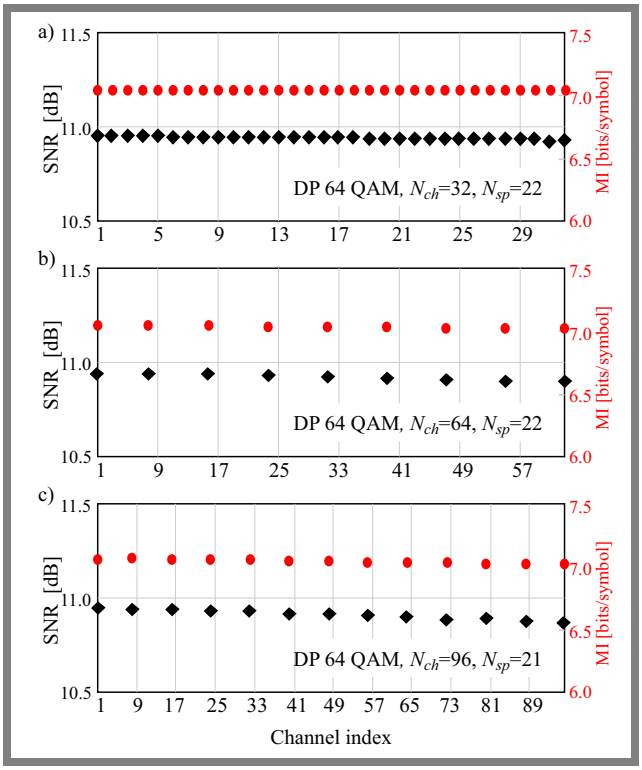

Fig. 9. SNR and MI with channel index for DP 64-QAM in WDM system operating at $P_L = -1$ dBm, and $R_s = 40$ Gbps: a) $N_{ch} = 32$, b) $N_{ch} = 64$, and c) $N_{ch} = 96$ with $N_{sp} = 22, 22, 21$, respectively.

3.2. Power Considerations

The performance of a DP 64-QAM PCS-WDM system is investigated with N_{ch} values of 32, 64, and 96, after a transmission over 20 spans, assuming $R_s = 40$ Gbps. To train the AE efficiently, it is essential to determine the range of P_L that yields favorable performance metrics. This range should

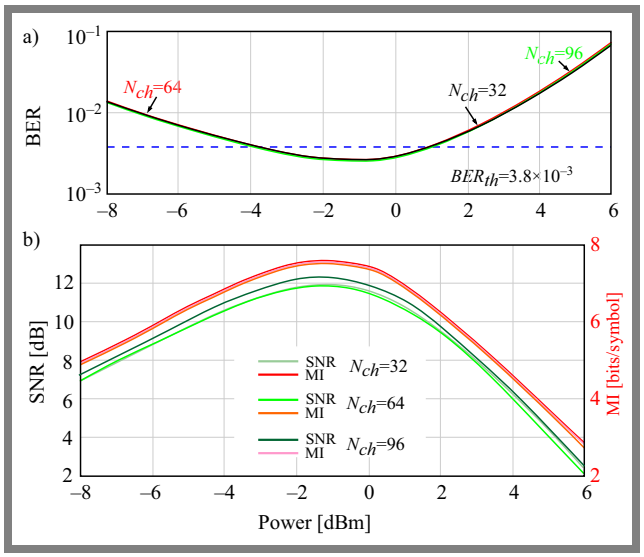

Fig. 10. Dependence of SNR, MI, and BER on channel launch power for a DP-QAM PCS-WDM system operating with $N_{sp} = 20$, and $R_s = 40$ Gbps.

guarantee that the AE is instructed to use launch powers that yield SNR and MI values within the dedicated power range by satisfying the BER requirement of 7% HD FEC or less. Figure 10 shows the variation of SNR, MI and BER with channel launch power P_L and taking N_{ch} as an independent parameter. The figure contains six parts, with parts a–b, c–d, and e–f corresponding to $N_{ch} = 32, 64$, and 96, respectively. The AE is trained at various power levels P_L within the –8 to 6 dBm range individually.

This work trains the AE with varying launch powers P_L and the number of channels N_{ch}, in order to achieve better SNR and MI ratios within a specific range of launch powers that renders BER below the threshold of 3.8×10^{-3}. In all instances where the AE is trained with $N_{ch} = 32, 64$, and 96, it is found that the optimum launch power range, which yields the highest SNR and MI values, is between –4 and 1 dBm.

Subsequently, the SNR and MI ratios are computed for launch power values of –4, –1, and 1 dBm for $N_{ch} = 32, 64$, and 96. These ratios yield optimal values of SNR and MI when the AE is trained within this specific range of launch power. The computed ratios of SNR are 0.90 and 0.91, and for MI, are 0.90 and 0.92 when $N_{ch} = 32$ is utilized to train the AE. However, for $N_{ch} = 64$ and 96, the computed ratio is similar and equals 0.90 for SNR in all instances, while for MI is 0.90 and 0.91.

Based on our findings, it appears that the developed AE is consistent and robust against different numbers of WDM channels utilized for training the AE-based PCS within a specific launch power range that satisfies the BER requirement of BER_{th} (i.e. 7% HD FEC). Based on the training results depicted in Fig. 10, it is essential to note that during the AE training phase, the highest possible SNR and MI values with the lowest possible BER are obtained at a P_L –1 dBm. Consequently, –1 dBm is regarded as the launch power for training the AE in this work.

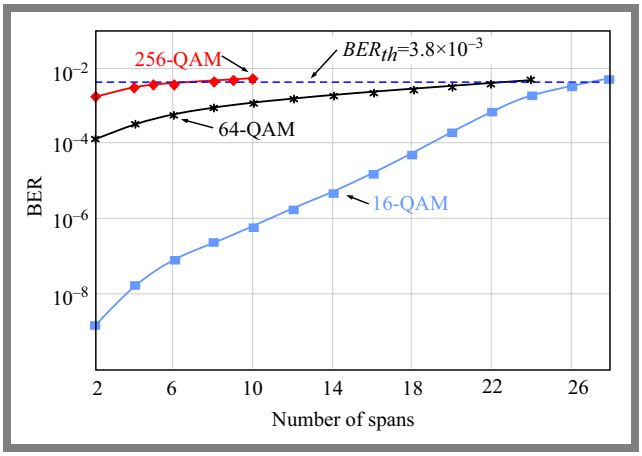

Fig. 11. Impact of transmission distance and DP modulation formats on BER of an AE-based PCS-WDM system operating with $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm.

3.3. Impact of Modulation Formats and Symbol Rate on Performance

The impact of modulation format and symbol rate on the transmission efficacy of an AE-based coherent PCS-WDM system is examined. Consideration is given to three DP modulation formats (16-, 64-, and 256-QAM) and four R_s values (20, 40, 60, and 80 Gbps) with channel spacing Δf set to 25, 50, 75, and 100 GHz, respectively. It is assumed that the parameters of each channel's AE are identical, and these parameters are acquired through the training process of the central channel's AE. The training procedure is executed for every pair of R_s and modulation formats.

The relationship between transmission distance and the corresponding received channel BER for various modulation formats is presented in Fig. 11 for $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm. This figure is useful for determining the MTR where BER is maintained at or below BER_{th} =3.8×10^{-3}. The results demonstrate that the MTR differs depending on the modulation format. For a 16-QAM format, the MTR is 26 spans. In contrast, the MTR diminishes as the modulation order goes up. Beyond that, the MTR for the 64-QAM format is reduced to 22 spans, while it is reduced to 6 spans for the 256-QAM format. The presented results indicate that an increase in modulation order leads to a reduction in the maximum distance.

The results shown in Fig. 12 illustrate how the MTR of 32-, 64-, and 96-channel PCS-WDM systems with DP 64- and 256-QAM signaling and a P_L of –1 dBm depends on the symbol rate. Training is conducted at four different values of R_s: 20, 40, 60, and 80 Gbps, with frequency channel spacings Δf of 25, 50, 75, and 100 GHz, respectively.

The objective is to ascertain the MTR at which the AE produces an appropriate response for every value of R_s. In the case of DP 64-QAM and 256-QAM signaling, the inquiry demonstrates that AE training using an $R_s = 20$ Gbps and $N_{ch} = 32, 64$, and 96 channels achieves MTR of 28, 28, and 27 spans for DP 64-QAM signaling, respectively, while DP 256-QAM achieves six spans for $N_{ch} = 32, 64$, and 96 channels. Recall that each span corresponds to a distance of

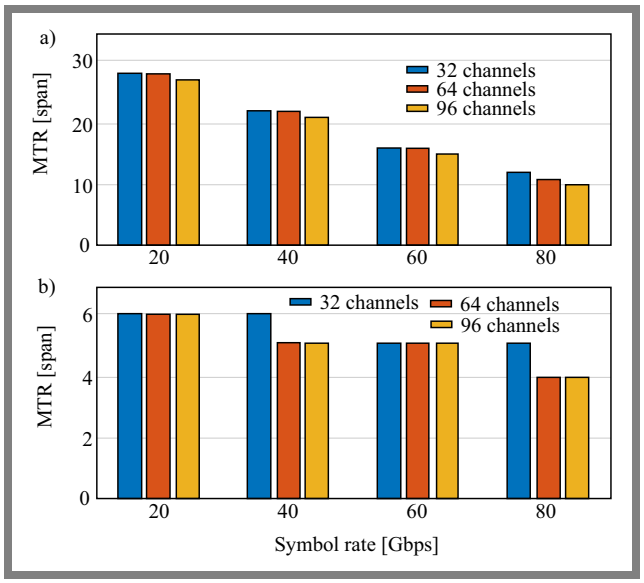

Fig. 12. Predictions of MTR for an AE-based WDM system designed with four values of symbol rate, using $N_{ch} = 32, 64$, and 96 with $P_L = -1$ dBm: a) DP 64-QAM and b) DP 256-QAM.

100 kilometers. Similarly, at R_s of 40 Gbps, for DP 64-QAM signaling, the MTR is 22, 22, and 21 spans for $N_{ch} = 32, 64$, and 96 channels, respectively.

In contrast, DP 256-QAM yields 6, 5, and 5 spans. At R_s 60 Gbps, the MTR reduces to 16, 16, and 15 spans for DP 64-QAM signaling, and DP 256-QAM achieves five spans for $N_{ch} = 32, 64$, and 96 channels. Finally, at R_s 80 Gbps, the MTR reduces to 12, 11, and 10 spans for DP 64-QAM signaling, and DP 256-QAM achieves 5, 4, and 4 spans for $N_{ch} = 32, 64$, and 96 channels, respectively. A comparative analysis of the performance of the AE when trained with identical system parameters as stated previously is presented in Tab. 2 detailing SNR, MI, and BER for various values of R_s.

The findings indicate that the response of the AE to using two modulation formats (DP 64- and 256-QAM) remains consistent when considering varied maximum reach distances and R_s values. This is achieved due to the method used (i.e., AE), involving integrating the transmitter (TX), the fiber channel, and the receiver (RX) into a single neural network (NN) supported by PCS. Then, the AE is jointly trained to replicate TX inputs using RX outputs. Thus, the structure of the AE is both flexible and consistent.

3.4. Performance Comparison

This subsection presents a performance comparison between the developed AE-based PCS-WDM system and a conventional system, with respect to Shannon limits. The performance of both systems is evaluated in terms of BER as a function of SNR. Furthermore, the comparison is extended to cover the dependence of SNR and MI with the number of spans.

The Shannon capacity theorem establishes the upper limit on the quantity of data that can be transmitted through a given medium or channel [36]:

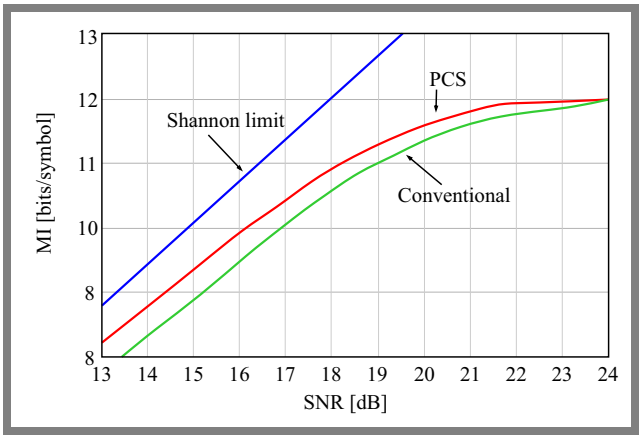

Fig. 13. Variation of MI with SNR for AE-based PCS-WDM and conventional systems, assuming $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm, for DP 64-QAM

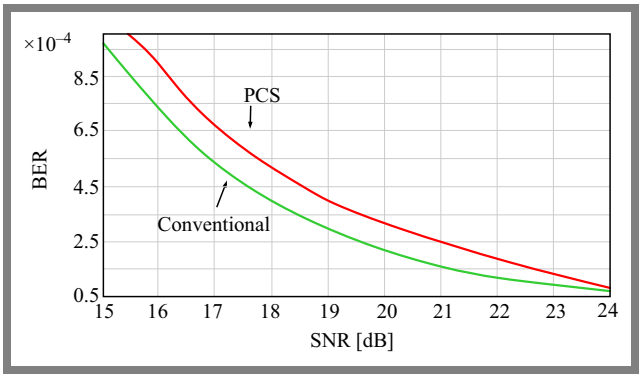

Fig. 14. BER with SNR for AE-based PCS-WDM and conventional systems, assuming $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm for DP 64-QAM.

$$C = B \log_2(1 + \frac{S}{N}), \tag{15}$$

where C is the channel's capacity in [bps], B represents the bandwidth available for data transmission in [Hz], s is the detected signal power, and the total channel noise power across bandwidth B is denoted by N.

Figure 13 shows the dependence of MI on SNR for both systems, as a function of the number of link spans. The parameters used in the investigation are DP 64-QAM signaling with $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm. The developed AE-based WDM system outperforms the conventional solution. The values of MI of the conventional and AE systems at SNR = 18 dB are 10.57 and 10.94 bits/symbol, respectively. This provides a 0.4 shaping gain. Thus, the performance of the developed AE-based PCS-WDM is better.

Figure 14 shows a comparison between AE-based PCS-WDM and conventional systems in terms of BER and SNR, as a function of the number of link spans. BERs of conventional and AE-based WDM systems at SNR = 18 dB are 3.75×10^{-4} and 3.25×10^{-4}, respectively. This provides an improvement in BER 0.5×10^{-4} and this enhancement signifies that the developed AE offers performance gains compared to the conventional system.

Tab. 2. Comparison of SNR, MI, and BER for AE-based PCS-WDM.

Modulation	Performance	Symbol rate [Gbps]			
		20	40	60	80
DP 64-QAM, $N_{ch} = 32$	SNR [dB]	10.94	10.95	10.94	11.05
	MI [bits/sym]	7.04	7.04	7.04	7.09
	BER [$\times 10^{-3}$]	3.62	3.61	3.62	3.64
DP 256-QAM, $N_{ch} = 32$	SNR [dB]	18.24	16.70	16.02	14.86
	MI [bits/sym]	11.42	10.48	10.07	9.37
	BER [$\times 10^{-3}$]	3.75	3.79	3.76	3.79
DP 64-QAM, $N_{ch} = 64$	SNR [dB]	10.83	10.92	10.93	11.15
	MI [bits/sym]	6.97	7.03	7.03	7.29
	BER [$\times 10^{-3}$]	3.75	3.64	3.63	3.58
DP 256-QAM, $N_{ch} = 64$	SNR [dB]	18.13	17.48	16.01	15.83
	MI [bits/sym]	11.37	10.97	10.07	9.96
	BER [$\times 10^{-3}$]	3.77	3.73	3.78	3.57
DP 64-QAM, $N_{ch} = 96$	SNR [dB]	10.94	11.31	11.01	11.32
	MI [bits/sym]	7.06	7.56	7.12	7.24
	BER [$\times 10^{-3}$]	3.61	3.57	3.58	3.62
DP 256-QAM, $N_{ch} = 96$	SNR [dB]	18.07	17.47	16.03	15.83
	MI [bits/sym]	11.33	10.96	10.07	9.96
	BER [$\times 10^{-3}$]	3.77	3.73	3.75	3.78

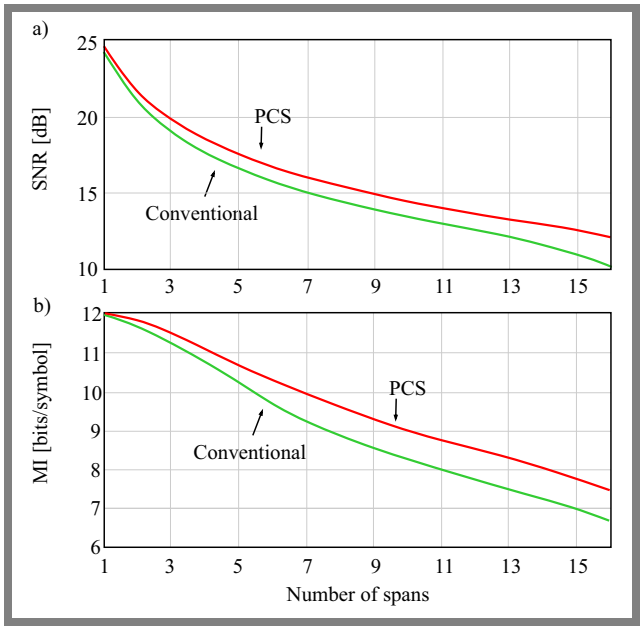

Fig. 15. Variation of: a) SNR and b) MI with the number of spans for AE-based PCS-WDM and conventional systems, for $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm, for DP 64-QAM.

Figure 15a-b shows a comparison between the conventional system and the AE-based PCS-WDM system at $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm in terms of SNR and MI as a function of the number of link spans, respectively. It is evident from this figure that the SNR and MI of the two systems are comparable for one and two spans. The developed AE-based PCS-WDM system exhibits better SNR and MI performance compared to the conventional system, as the number of spans increases.

3.5. Constellation Comparison of AE for Different Number of Spans

Figure 16 displays the learned constellation at different numbers of spans for the developed AE-based PCS-WDM system, using $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm for DP 64-QAM. The color of the constellation points represents their occurrence probabilities and the color map – ranging from blue to red – represents values ranging from low to high.

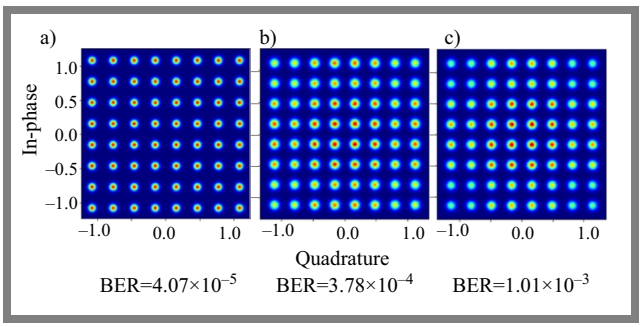

Fig. 16. Learned constellation at: a) 1, b) 5, and c) 10 spans for AE-based PCS-WDM system, assuming $N_{ch} = 32$, $R_s = 40$ Gbps, and $P_L = -1$ dBm for DP 64-QAM.

Red represents the highest probability of occurrence, and blue is the lowest probability of occurrence.

At one span, the learned constellation demonstrates a high probability for each of its points, in addition to a uniform distribution and a low BER of 4.07×10^{-5}. When the number of spans is increased to five and ten, however, BER increases to 3.78×10^{-4}, and 1.01×10^{-3}, respectively. Based on these results, it can be inferred that as the number of spans increases, the occurrence probabilities of points within the learned constellation diminish, leading to a greater BER. Furthermore, it has been observed that the constellation that has been learned to tolerate NLI noise.

4. Conclusion

In the course of the work described in this paper, an E2EDL-based PCS for M-QAM signaling (M=16, 64 and 256) that minimizes the effects of nonlinear fiber optics while utilizing a high-capacity coherent WDM system has been developed. The simulation results show that optimal system performance can be obtained when the AE consists of 2 hidden layers, 32 neurons per hidden layer, and a "$32 \times M$" batch size to obtain. The simulation results show that training the AE of each channel for DP 64-QAM, $N_{ch} = 32, 64$, and 96, with $N_{sp} = 22, 22, 21$, respectively, and $P_L = -1$ dBm gives approximately the same performance predictions compared to the case when all the channels use the same central channel trained AE. This observation demonstrates the consistency and dependability of the developed AE. Training has been conducted at four different values of R_s: 20, 40, 60, and 80 Gbps, with frequency channel spacings Δf of 25, 50, 75, and 100 GHz. The objective is to ascertain the MTR at which the AE produces an appropriate response for every value of R_s.

The AE has been trained with different launch power P_L values and the numbers of channels N_{ch} to assess SNR and MI within a specific range of P_L that renders a below-threshold BER. For $N_{ch} = 32, 64$, and 96, the P_L range is from –4 to 1 dBm, and the optimum P_L value that yields a minimum BER is –1 dBm. SNR and MI reach their maximum levels at the optimum value of P_L and reduce to about 90% when $P_L = -4$ and 1 dBm. The developed AE-based PCS-WDM provides a 0.4 shaping gain and outperforms the conventional system. In the future, additional progress is expected with adequate extensions of the system. It is important to assess the efficacy of the developed AE regarding other NLIN models or use other fiber models. In addition, it enhances system performance by joint geometric constellation shaping and PCS to optimize both the position and probability of the symbols in the constellation.

Acknowledgments

The researchers express their gratitude to the College of Engineering at Al-Nahrain University for supplying the necessary resources to carry out this study. Mrs. Ayam is appreciative that Mustansirityah University's College of Engineering has awarded her the Ph.D. scholarship.

References

[1] X. Jia *et al.*, "Nonlinear Transmission Performance Comparison Employing 60×416.7-Gb/s TPS-64QAM and UD-16QAM Systems with Limited Bandwidth", *Optics Letters*, vol. 48, no. 19, pp. 5169–5172, 2023 (https://doi.org/10.1364/OL.503914).

[2] A. Kaushik and H. Saini, "Evaluation of WDM System Performance over a Range of Input System Design Parameters", *Journal of Optics*, vol. 52, no. 3, pp. 1148–1153, 2023 (https://doi.org/10.1007/s12596-022-00949-2).

[3] A.A. Salman, G.B. Esmer, M.H. Ali, and W.K. Al-Azzawi, "Design and Simulation of 40 GHz-WDM Communication System-based Optical Frequency Comb Generator", *Journal of Optics*, vol. 53, pp. 538–543, 2023 (https://doi.org/10.1007/s12596-023-01177-y).

[4] X. Jia *et al.*, "Many-to-one Mapping Based Truncated Probabilistic Shaped 64-QAM for Metro Transmissions Enabled by Polar Code", *Journal of Lightwave Technology*, vol. 41, no. 21, pp. 6700–6709, 2023 (https://doi.org/10.1109/JLT.2023.3290292).

[5] Z. Liu *et al.*, "Achievable Information Rate Optimization in C-band Optical Fiber Communication System", *Frontiers of Optoelectronics*, vol. 16, pp. 17–22, 2023 (https://doi.org/10.1007/s12200-023-00072-5).

[6] A. Rode, B. Geiger, S. Chimmalgi, and L. Schmalen, "End-to-end Optimization of Constellation Shaping for Wiener Phase Noise Channels with a Differentiable Blind Phase Search", *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3849–3859, 2023 (https://doi.org/10.1109/JLT.2023.3265308).

[7] P. Zou, F. Hu, Y. Zhao, and N. Chi, "On the Achievable Information Rate of Probabilistic Shaping QAM Order and Source Entropy in Visible Light Communication Systems", *Applied Sciences*, vol. 10, no. 12, pp. 4299–4309, 2020 (https://doi.org/10.3390/app10124299).

[8] D. Che, J. Cho, and X. Chen, "Does Probabilistic Constellation Shaping Benefit IM-DD Systems Without Optical Amplifiers?", *Journal of Lightwave Technology*, vol. 39, no. 15, pp. 4997–5007, 2021 (https://doi.org/10.1109/JLT.2021.3083530).

[9] Z. Wang *et al.*, "Probabilistic Shaping Based Constellation Encryption for Physical Layer Security in OFDM RoF System", *Optics Express*, vol. 29, no. 12, pp. 17890–17901, 2021 (https://doi.org/10.1364/OE.424661).

[10] B. Oliveira *et al.*, "Optimizing Probabilistic Constellation Shaping for Amplifier-less Coherent Optical Links", *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4318–4330, 2021 (https://doi.org/10.1109/JLT.2021.3072547).

[11] S. Zhou *et al.*, "A High-security Probabilistic Constellation Shaping Transmission Scheme Based on Recurrent Neural Networks", *Photonics*, vol. 10, no. 10, art. no. 1078, 2023 (https://doi.org/10.3390/photonics10101078).

[12] M.A. Amirabadi, M.H. Kahaei, and S.A. Nezamalhosseini, "End-to-end Deep Learning for Joint Geometric-probabilistic Constellation Shaping in FMF System", *Physical Communication*, vol. 55, art. no. 101903, 2022 (https://doi.org/10.1016/j.phycom.2022.101903).

[13] W. Jiang *et al.*, "End-to-end Learning of Constellation Shaping for Optical Fiber Communication Systems", *IEEE Photonics Journal*, vol. 15, no. 6, pp. 1–7, 2023 (https://doi.org/10.1109/JPHOT.2023.3321736).

[14] S. Cammerer *et al.*, "Trainable Communication Systems: Concepts and Prototype", *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5489–5503, 2020 (https://doi.org/10.1109/TCOMM.2020.3002915).

[15] P. Ge *et al.*, "Autoencoder Assisted Subcarrier Optimization for Nonlinear Frequency Division Multiplexing", *Optics Express*, vol. 31, no. 21, pp. 34443–34458, 2023 (https://doi.org/10.1364/oe.500473).

[16] B. Karanov, P. Bayvel, and L. Schmalen, "End-to-end Learning in Optical Fiber Communications: Concept and Transceiver Design", *2020 European Conference on Optical Communications (ECOC)*, Brussels, Belgium, 2020 (https://doi.org/10.1109/ECOC48923.2020.9333174).

[17] T. Uhlemann *et al.*, "Deep-learning Autoencoder for Coherent and Nonlinear Optical Communication", *Photonic Networks; 21th ITG-Symposium*, Online, 2020 (https://doi.org/10.48550/arXiv.2006.15027).

[18] S. Gaiarin, F. Da Ros, R.T. Jones, and D. Zibar, "End-to-end Optimization of Coherent Optical Communications over the Split-step Fourier Method Guided by the Nonlinear Fourier Transform Theory", *Journal of Lightwave Technology*, vol. 39, no. 2, pp. 418–428, 2021 (https://doi.org/10.1109/JLT.2020.3033624).

[19] Z. Niu *et al.*, "End-to-end Deep Learning for Long-haul Fiber Transmission using Differentiable Surrogate Channel", *Journal of Lightwave Technology*, vol. 40, no. 9, pp. 2807–2822, 2022 (https://doi.org/10.1109/JLT.2022.3148270).

[20] V. Oliari *et al.*, "High-cardinality Hybrid Shaping for 4D Modulation Formats in Optical Communications Optimized via End-to-end Learning", 2021 (http://arxiv.org/abs/2112.10471).

[21] M. Stark, F.A. Aoudia, and J. Hoydis, "Joint Learning of Geometric and Probabilistic Constellation Shaping", 2019 (http://arxiv.org/abs/1906.07748).

[22] V. Aref and M. Chagnon, "End-to-end Learning of Joint Geometric and Probabilistic Constellation Shaping", *Optical Fiber Communications Conference (OFC)*, San Diego, USA, 2022 (https://doi.org/10.1364/OFC.2022.W4I.3).

[23] X. Liu, I. Darwazeh, N. Zein, and E. Sasaki, "Probabilistic Shaping for Multidimensional Signals with Autoencoder-based End-to-end Learning", *IEEE Wireless Communications and Networking Conference (WCNC)*, Austin, USA, 2022 (https://doi.org/10.1109/WCNC51071.2022.9771910).

[24] F. Wang, G. Hu, and Z. Li, "A Novel Four Dimensional Constellation Shaping with Non-uniform Signaling for Long-haul Fiber-optic Communication", *Optics Communications*, vol. 486, art. no. 126755, pp. 1-6, 2021 (https://doi.org/10.1016/j.optcom.2021.126755).

[25] B. Chen *et al.*, "Geometrically-shaped Multidimensional Modulation Formats in Coherent Optical Transmission Systems", *Journal of Lightwave Technology*, vol. 41, no. 3, pp. 897–910, 2023 (https://doi.org/10.1109/JLT.2022.3204101).

[26] O. Jovanovic, F. Da Ros, D. Zibar, and M.P. Yankov, "Geometric Constellation Shaping for Fiber-optic Channels via End-to-end Learning", *Journal of Lightwave Technology*, vol. 41, no. 12, pp. 3726–3736, 2023 (https://doi.org/10.1109/JLT.2023.3276300).

[27] K. Suetake, S. Ikegawa, R. Saiin, and Y. Sawada, "S^3NN: Time Step Reduction of Spiking Surrogate Gradients for Training Energy Efficient Single-step Spiking Neural Networks", *Neural Networks*, vol. 159, pp. 208–219, 2023 (https://doi.org/10.1016/j.neunet.2022.12.008).

[28] E. Jang, S. Gu, and B. Poole, "Categorical Reparameterization with Gumbel-Softmax", *5th International Conference on Learning Representations (ICLR)*, Toulon, France, 2017 (https://doi.org/10.48550/arXiv.1611.01144).

[29] I.A.M. Huijben, W. Kool, M.B. Paulus, and R.J.G. van Sloun, "A Review of the Gumbel-max Trick and its Extensions for Discrete Stochasticity in Machine Learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1353–1371, 2023 (https://doi.org/10.1109/TPAMI.2022.3157042).

[30] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Properties of Nonlinear Noise in Long, Dispersion-Uncompensated Fiber Links", *Optics Express*, vol. 21, no. 22, pp. 25685–25699, 2013 (https://doi.org/10.1364/OE.21.025685).

[31] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Accumulation of Nonlinear Interference Noise in Fiber-Optic Systems", *Optics Express*, vol. 22, no. 12, pp. 14199–14211, 2014 (https://doi.org/10.1364/OE.22.014199).

[32] O. Jovanovic, M.P. Yankov, F. Da Ros, and D. Zibar, "End-to-end Learning of a Constellation Shape Robust to Channel Condition Uncertainties", *Journal of Lightwave Technology*, vol. 40, no. 10, pp. 3316–3324, 2022 (https://doi.org/10.1109/JLT.2022.3169993).

[33] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Polarization-Ring-Switching for Nonlinearity-Tolerant Geometrically Shaped Four-Dimensional Formats Maximizing Generalized Mutual Information", *Journal of Lightwave Technology*, vol. 37, no. 14, pp. 3579–3591, 2019 (https://doi.org/10.1109/JLT.2019.2918072).

[34] R.A. Shafik, M.S. Rahman, and A.R. Islam, "On the Extended Relationships Among EVM, BER and SNR as Performance Metrics", *International Conference on Electrical and Computer Engineering*, Dhaka, Bangladesh, 2006 (https://doi.org/10.1109/ICECE.2006.355657).

[35] G.P. Agrwal, *Fiber-Optic Communication Systems*, Wiley, 5th ed., 2021 (ISBN: 9781119737360).

[36] M. Parker, "Digital Signal Processing 101: Everything You Need to Know to Get Started", Elsevier, 2nd ed., 2017 (ISBN: 9780128114537).

Ayam M. Abbass, M.Sc.
Department of Laser and Optoelectronics Engineering,
College of Engineering
https://orcid.org/0000-0003-3101-6761
E-mail: ayammohsen@uomustansiriyah.edu.iq
Al-Nahrain University, Baghdad, Iraq
https://nahrainuniv.edu.iq
Mustansiriyah University, Baghdad, Iraq
https://uomustansiriyah.edu.iq

Raad Sami Fyath, Ph.D., Professor
Department of Computer Engineering,
College of Engineering
https://orcid.org/0000-0002-1029-3471
E-mail: raad.s.fyath@nahrainuniv.edu.iq
Al-Nahrain University, Baghdad, Iraq
https://nahrainuniv.edu.iq