Physical Communication 55 (2022) 101903

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

End-to-end deep learning for joint geometric-probabilistic constellation shaping in FMF system

Mohammad Ali Amirabadi, Mohammad Hossein Kahaei, S. Alireza Nezamalhosseini *

School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran, 1684613114, Iran

ARTICLE INFO

Article history:
Received 20 January 2022
Received in revised form 3 August 2022
Accepted 21 September 2022
Available online 4 October 2022

Keywords:
End-to-end deep learning
Joint geometric-probabilistic constellation shaping
Few-mode fiber nonlinearity

ABSTRACT

The few-mode fiber (FMF) nonlinear effects including Kerr nonlinearity and nonlinear coupling are the most important barriers in FMF-based transmission. The performance degradation due to FMF nonlinearity can be mitigated by properly designing the constellation points. The location and occurrence probability of constellation points can be optimized with geometric constellation shaping (GCS) and probabilistic constellation shaping (PCS), respectively. In this paper, we present end-to-end deep learning (EEDL)-based GCS, PCS, and joint geometric-probabilistic constellation Shaping (JGPCS) algorithms for FMF systems. The performance of the proposed algorithms is compared with the uniform distributed quadrature amplitude modulation (QAM) and the well-known Maxwell–Boltzmann distributed QAM, in terms of mutual information (MI). Simulation results show 0.15, 0.19, and 0.22 *bits/symbol* MI improvement respectively for proposed EEDL-based GCS, PCS, and JGPCS algorithms compared with uniform distributed QAM constellation.

1. Introduction

Over the last decades, the required capacity in the optical fiber communication systems has been increased exponentially. Due to the ever increasing traffic generation, the single mode fiber (SMF) is rapidly approaching its capacity limits [1]. Consequently, different multiplexing schemes have been developed to increase the optical fiber communication capacity, among them, the mode-division multiplexing is a very promising solution wherein different parallel data streams can be transmitted over spatial modes in a few-mode fiber (FMF) link [2]. The FMF nonlinear effects, including Kerr nonlinearity and nonlinear coupling, are the main barriers in front of FMF practical applications [3–6]. Optimizing the location of the constellation points, i.e., geometric constellation shaping (GCS), or the probabilities of occurrence of the constellation points, i.e., probabilistic constellation shaping (PCS) can mitigate the fiber nonlinearity and improve the system performance in terms of mutual information (MI) [7]. The GCS is employed in shape of the multi-ring constellations in [8] and iterative polar modulation in [9,10] to arrange the location of constellation point to approximate a Gaussian distribution. GCS produces irregular constellation points which increases the receiver complexity, and does not provide a simple solution for designing the location of constellation points in any channel conditions [11]. PCS optimizes the occurrence probability of constellation points for quadrature amplitude modulation (QAM) to approximate the Gaussian signaling [11]. It is shown that in the additive white Gaussian noise (AWGN) channel, the well-known Maxwell–Boltzmann distribution is the optimal PCS for QAM constellation [12]. The Maxwell–Boltzmann distributed QAM, although optimal, has some barriers while practical implementation. In this distribution, the inner constellation points appear more likely rather than the outer points, therefore, a very high number of symbols is necessary for precisely approximating the desired distribution. The digital to analog converter (DAC) restricts the transmit sequence length in practical implementations [13], thereby, it is extremely hard to build a transmit sequence with a desired Maxwell–Boltzmann distribution [14]. Combination of PCS and GCS in terms of joint geometric-probabilistic constellation shaping (JGPCS) is investigated in [15,16]. Although JGPCS shows MI gain over PCS, it is along with disadvantages of both GCS and PCS.

Recently, deep learning (DL) has attracted significant attentions in optical communication for fiber nonlinearity mitigation [17], modulation format identification [18], optical performance monitoring [19], and resource allocation [7,20]. Considering FMF communication systems, several studies have been proposed [21] which employ DL for demodulation [22], mode decomposition [23], detection [24], and equalization [25]. However, the aforementioned studies in FMF consider DL at the receiver side which result in sub-optimal solutions. Using the DL at both transmitter and receiver sides, i.e. end-to-end DL (EEDL), can enhance

* Corresponding author.
E-mail addresses: m_amirabadi@elec.iust.ac.ir (M.A. Amirabadi), kahaei@iust.ac.ir (M.H. Kahaei), nezam@iust.ac.ir (S.A. Nezamalhosseini).

https://doi.org/10.1016/j.phycom.2022.101903
1874-4907/

the performance significantly [26]. In EEDL approach, the transmitter, fiber channel, and receiver are implemented as an autoencoder, and the transceiver is jointly trained to best match the output with the input [27]. EEDL can design the constellation shape by optimizing either the location of the symbols or their occurrence probabilities. The optimum constellation for the FMF system is robust to FMF nonlinearity [28]. The self- and cross-channel as well as intra- and inter-modal FMF nonlinear interactions can be well modeled as an additive Gaussian noise source by enhanced Gaussian noise (EGN) model [4–6]. The EGN model formulates the relationship between nonlinear interference (NLI) noise variance and high-order moments of the transmitted signal constellation [27]. The NLI noise variance is small in constellations with small high-order moments [29], thereby an optimal set of high-order moments (an optimal constellation) exists which maximizes MI [30,31]. The constellation shaping is a non-convex optimization problem conventionally solved by metaheuristic algorithms such as genetic algorithm or iterative optimization methods which are time-consuming approaches [32–34]. The main feature of EEDL-based constellation shaping is that it unlocks using the gradient-based optimization techniques, and solves constellation shaping problem via gradient-descent techniques which are much more cost-effective than metaheuristic or iterative optimization approaches. The constellation shaping is analytically solved for simple scenarios such as AWGN channel in which channel distribution is known. However, in case of a mathematical intractability, either because of complexity or even lack of channel distribution it is impossible to mathematically design constellations, unless several assumptions are made which yields to sub-optimal solutions. Note that finding the optimal distribution using conventional PCS methods is a difficult problem even when knowing channel distribution. Moreover, conventional PCS methods assume that the target distribution is symmetric around the origin which is true for AWGN channel. However, the EEDL-based constellation shaping is not restricted to symmetric probability distributions, and can be employed for any channel model.

1.1. Literature review

The available EEDL-based constellation shaping investigations can be divided into three categories, GCS, PCS, and JGPCS. Considering GCS category, [35] employed GCS to learn a constellation robust to SNR and laser linewidth estimation errors with a MI gain up to 0.3 bits/symbol. [7,36–38] deployed GCS to optimize the constellation design mitigating nonlinear effects with gains up to 0.07 bits/symbol when trained with GN and NLI noise (NLIN) channel model. Final verification of these works is by split-step Fourier method (SSFM) simulations [7,16,36] and experimental setup [30,31]. EEDL is used in [37] for GCS considering a simplified single mode fiber channel mitigating nonlinear phase noise effect. In these works, almost the same deep neural network (DNN)-based modulators is considered, and only the learned constellation is used in transmitter after training. While dealing with GN and NLIN models, the channel model appears memoryless Gaussian auxiliary from the receiver standpoint and the DNN-based detector mimics maximum-likelihood receiver. Therefore, the complexity and performance of presented DNN-based detectors is the same. In fact, these works train almost the same EEDL-based algorithms considering different system and channel models, i.e., the methodology is the same but implementation is different. Considering PCS and JGPCS categories, PCS/JGPCS is its infancy in optical communication applications, an auxiliary AWGN channel (not GN or NLIN model) is considered in [39–41] showing about 0.04 bit/symbol MI improvement in JGPCS compared with PCS. In [39] a training symbol batch of size N is generated by drawing randomly N samples indices from p_m, then each index m is taken about Np_m times, for $m = \{1, \ldots, M\}$, and all indices are randomly permuted. In [40] the sampling is done by finding the maximizing argument of the summation a sample with Gumble distribution and $log_2(p_m)$. Both approaches show the same performance as Maxwell–Boltzmann distribution and their complexity is almost the same. However, [39] needs some careful integer rounding when $Np_m < 1$ and [40] needs to tune a hyperparameter.

1.2. Novelties and contributions

Despite the importance of constellation shaping, it is only investigated in SMF systems, and no conventional or EEDL-based approach is employed for GCS, PCS, or JGPCS in FMF systems. In this paper, we propose EEDL-based GCS, PCS, and JGPCS algorithms for mitigating FMF nonlinear effects. The contributions and novelties of this paper include;

- Employing constellation shaping in FMF system which can provide higher MI for FMF-based transmission,
- Deploying EEDL in FMF system that can optimize the FMF transceiver design in an end-to-end manner and obtain better results than disjoint block based processing,
- Presenting EEDL-based GCS, PCS, and JGPCS algorithms for FMF system with more flexibility and be more practical than the conventional analytical methods,
- Providing a comprehensive comparison by demonstrating well-known conventional methods and EEDL-based algorithms considering both EGN model and verifying the practicability of the obtained results by final verification based on the SSFM.

The rest of this paper is organized in the following form; the FMF system model and the signal propagation are described in Section 2. Section 3 explains the proposed EEDL-based algorithms, Section 4 provides the simulation results, and Section 5 is the discussion, and Section 6 presents the conclusion of this paper.

2. FMF system model

The transmitted signal is a multiplexing of 2 polarization modes, D spatial modes, and N_{ch} frequency channels. The considered FMF is consisted of N_s spans with a Few-Mode Erbium-Doped Fiber Amplifier (FM-EDFA) at the end of each span to compensate fiber attenuation. In turn, each FM-EDFA produces Amplified Spontaneous Emission (ASE) noise. FMF linear effects affecting signal propagation include modal dispersion, chromatic dispersion, and linear coupling [42]. Kerr-nonlinearity and nonlinear coupling are the considered FMF nonlinear effects [42]. The received signal is passed through a Multi-Input Multi-Output (MIMO) Digital Signal Processing (DSP) for compensating the FMF linear effects including modal dispersion, chromatic dispersion, and linear coupling. Then Carrier Phase Recovery (CPR) is deployed to recover the nonlinear phase noise. The received signal at the output of the CPR can be described by EGN model as a summation of the transmitted signal with ASE and NLI noise [4]. Therefore, considering $x_{n,m}$ as the transmitted symbol in nth channel and mth mode, the received signal, $y_{n,m}$, can be expressed as [4]

$$y_{n,m} = x_{n,m} + n_{ASE,n,m} + n_{EGN,n,m}, \tag{1}$$

where $n_{ASE,n,m} \sim N(0, \sigma^2_{ASE,n,m})$ and $n_{EGN,n,m} \sim N(0, \sigma^2_{EGN,n,m})$ respectively represent the ASE noise and NLI noise related to nth channel and mth mode. Here, $\sigma^2_{ASE,n,m} = N_s F_{n,m}(G_{n,m} - 1)h\nu\Delta f_{n,m}$ where $F_{n,m}$ is the amplifier noise figure of nth channel and mth

mode, N_s is number of spans, $G_{n,m}$ is amplifier gain of nth channel and mth mode equal to per span fiber loss of nth channel and mth mode, $\Delta f_{n,m}$ is the channel bandwidth of nth channel and mth mode, and h is the Planck's constant, and ν is central frequency [4]. The NLI variance of nth channel and mth mode, $\sigma^2_{EGN,n,m}$, can be defined as [4]

$$\begin{aligned}\sigma^2_{EGN,n,m} = \sum_{q=1}^{D}\Bigg[&\sum_{k_2,m_2,n_2} 3\kappa_1^{(k_2,q)}\kappa_1^{(m_2,q)}\kappa_1^{(n_2,m)}P_{k_2,q}P_{m_2,q}P_{n_2,m} \\ &\times X^a_{n,m}(k_2,m_2,n_2,q) + \sum_{k_2,n_2}\kappa_2^{(k_2,q)}\kappa_1^{(n_2,q)}5 \\ &\times (P^2_{k_2,q}P_{n_2,m}X^b_{n,m}(k_2,k_2,n_2,q) \\ &+ P_{k_2,m}P_{k_2,q}P_{n_2,q}X^c_{n,m}(k_2,n_2,k_2,q)) + \sum_{n_2}\kappa_3^{(n_2,q)}P^2_{n_2,q}P_{n_2,m} \\ &\times X^d_{n,m}(n_2,n_2,n_2,q)\Bigg],\end{aligned} \tag{2}$$

where

$$\begin{aligned}X^a_{n,m}(k_2,m_2,n_2,q) &= \frac{\tilde{\gamma}^2_{mq}}{4}\iiint_{-\infty}^{\infty}|\eta(f,f_1,f_2)|^2 g^{k_2,q} \\ &\times (f+f_1+f_2)g^{m_2,q}(f+f_2)g^{n_2,m}(f+f_1)g^{n,m}(f)df_1df_2df \\ X^b_{n,m}(k_2,k_2,n_2,q) &= \frac{\tilde{\gamma}^2_{mq}}{4}\iiint_{-\infty}^{\infty}|\eta(f,f_1,f_2)|^2 g^{k_2,q} \\ &\times (f+f_1+f_2)g^{k_2,q}(f+f_2)g^{n_2,m}(f+f_1)g^{n,m}(f)df_1df_2df \\ X^c_{n,m}(k_2,n_2,k_2,q) &= \frac{\tilde{\gamma}^2_{mq}}{4}\iiint_{-\infty}^{\infty}|\eta(f,f_1,f_2)|^2 g^{n_2,q} \\ &\times (f+f_1+f_2)g^{k_2,q}(f+f_2)g^{k_2,m}(f+f_1)g^{n,m}(f)df_1df_2df \\ X^d_{n,m}(n_2,n_2,n_2,q) &= \frac{\tilde{\gamma}^2_{mq}}{4}\iiint_{-\infty}^{\infty}|\eta(f,f_1,f_2)|^2 g^{n_2,q} \\ &\times (f+f_1+f_2)g^{n_2,q}(f+f_2)g^{n_2,m}(f+f_1)g^{n,m}(f)df_1df_2df,\end{aligned} \tag{3}$$

with

$$\begin{aligned}\eta(f,f_1,f_2) = &\frac{1-e^{-\alpha_{n,m}L_s}e^{j|\beta_q(f+f_1-f_2)-\beta_m(f)+\beta_m(f_2)-\beta_q(f_1)|L_s}}{\alpha_{n,m}-j|\beta_q(f+f_1-f_2)-\beta_m(f)+\beta_m(f_2)-\beta_q(f_1)|} \\ &\times\frac{1-e^{j|\beta_q(f+f_1-f_2)-\beta_m(f)+\beta_m(f_2)-\beta_q(f_1)|N_sL_s}}{1-e^{j|\beta_q(f+f_1-f_2)-\beta_m(f)+\beta_m(f_2)-\beta_q(f_1)|L_s}}.\end{aligned} \tag{4}$$

where $\kappa_1^{(n,m)} = \mu_2^{(n,m)}$, $\kappa_2^{(n,m)} = \mu_4^{(n,m)} - 2\mu_2^{(n,m)^2}$, and $\kappa_3^{(n,m)} = \mu_6^{(n,m)} - 4\mu_4^{(n,m)}\mu_2^{(n,m)} + 12\mu_2^{(n,m)^3}$, with $\mu_2^{(n,m)}$, $\mu_4^{(n,m)}$, and $\mu_6^{(n,m)}$ denoting the second, fourth, and sixth order moments of the constellation of nth channel and mth mode, respectively. Here, $P_{n,m}$ and $g^{(n,m)}(.)$ are respectively the launched power and the spectral shape of transmitted signal in nth channel and mth mode, $\alpha_{n,m}$ is attenuation of nth and mth mode and L_s is span length, and β_{1m} and β_{2m} are respectively the modal and chromatic dispersion coefficients of mth mode [4].

$$\tilde{\gamma}_{mq} = \begin{cases}\frac{4}{3}(\frac{2}{3})^{\delta_{mq}}f_{mq}\gamma & weak\ coupling \\ \kappa\gamma & strong\ coupling,\end{cases} \tag{5}$$

with

$$\kappa = \sum_{\substack{q\le p \\ m,q\in\{1,2,\ldots,D\}}}\frac{32}{2^{\delta_{mq}}}\frac{f_{mq}}{6D(2D+1)}, \tag{6}$$

where γ is the Kerr nonlinearity coefficient, $f_{mq} = \frac{A_{eff}}{I_mI_q}\iint F^2_m(x,y)F^2_q(x,y)dxdy$ is the nonlinear coupling coefficient between modes p and q, and A_{eff} is the effective area of the fundamental mode.

3. Proposed EEDL-based algorithms

The proposed EEDL-based GCS, PCS, and JGPCS algorithms for FMF system are depicted by Fig. 1a, b, and c, respectively, and we describe each of them in the following sections.

3.1. GCS

The proposed EEDL-based GCS algorithm demonstrated by Fig. 1a is composed of two DNN-based functions representing modulator and detector, with the goal of reproducing the generated symbol at the detector output. To do so, we first convert the generated M-ary symbol of nth channel and mth mode to the one-hot vector $\boldsymbol{s_{n,m}} \in S = \{e_i | i = 1,\ldots,M\}$ where e_i is equal to 1 at ith element and 0 elsewhere and M is modulation order (e.g. e_3 is [0, 0, 1, 0] when $M = 4$). Then $\boldsymbol{s_{n,m}}$ is modulated to constellation point $x_{n,m}$ via the following function

$$x_{n,m} = f_{\boldsymbol{\theta_M}}(\boldsymbol{s_{n,m}}), \tag{7}$$

where $f_{\boldsymbol{\theta_M}}(.)$ is DNN-based modulator with $\boldsymbol{\theta_M}$ as trainable parameters. The DNN-based modulator is composed of an input layer with M neurons, an output layer with 2 neurons, and N_{hid} hidden layers which each of them has N_{neu} hidden neurons. The output neurons depict in-phase and quadrature components of the transmitted signal. The DNN-based modulator is ended by a normalization layer to justify the unit power constraint of constellation design. The modulated symbol, $x_{n,m}$, is propagated through FMF nonlinear channel according to the following expression

$$y_{n,m} = f_{EGN}(x_{n,m}), \tag{8}$$

where $y_{n,m}$ is the received symbol of nth channel and mth mode at the output of CPR block. The received symbol is passed through the DNN-based detector according to the following formulation

$$\boldsymbol{r_{n,m}} = f_{\boldsymbol{\theta_D}}(y_{n,m}), \tag{9}$$

where $\boldsymbol{r_{n,m}}$ is the detected symbol of nth channel and mth mode, $f_{\boldsymbol{\theta_D}}(.)$ is the DNN-based detector with trainable parameters $\boldsymbol{\theta_D}$. The DNN-based detector is composed of an input layer with 2 neurons, an output layer with M neurons, and N_{hid} hidden layers which each of them has N_{neu} hidden neurons. The output neurons depict the demodulated symbol in one-hot vector form. The aim is to adjust $\boldsymbol{\theta} = \{\boldsymbol{\theta_M}, \boldsymbol{\theta_D}\}$ by end-to-end training the DNN-based modulator and DNN-based detector such that we have the minimum detection error. To do so, we use softmax activation function layer at the DNN output which yields the following probability (note that the softmax output can be seen as probability) [43]

$$\boldsymbol{r_{n,m}} \in \{p \in R^+_M | \sum_{i=1}^{M} p_i = 1\}, \tag{10}$$

then, we consider cross-entropy loss function which can be expressed as [44]

$$L(\boldsymbol{\theta}) = -1/K\sum_{k=1}^{K}\sum_{i=1}^{M}s^{(i,k)}_{n,m}log(r^{(i,k)}_{n,m}), \tag{11}$$

where K defines the considered batch size and the batch number is denoted by superscript k. The loss function can be iteratively minimized by stochastic gradient descent (SGD) method using the following formula

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} - \eta\nabla_{\boldsymbol{\theta}}\tilde{L}(\boldsymbol{\theta}^{(j)}), \tag{12}$$

where η is the learning rate, superscript j defines the iteration number, and the estimated gradient of loss function is denoted by $\nabla_{\boldsymbol{\theta}}\tilde{L}(.)$.

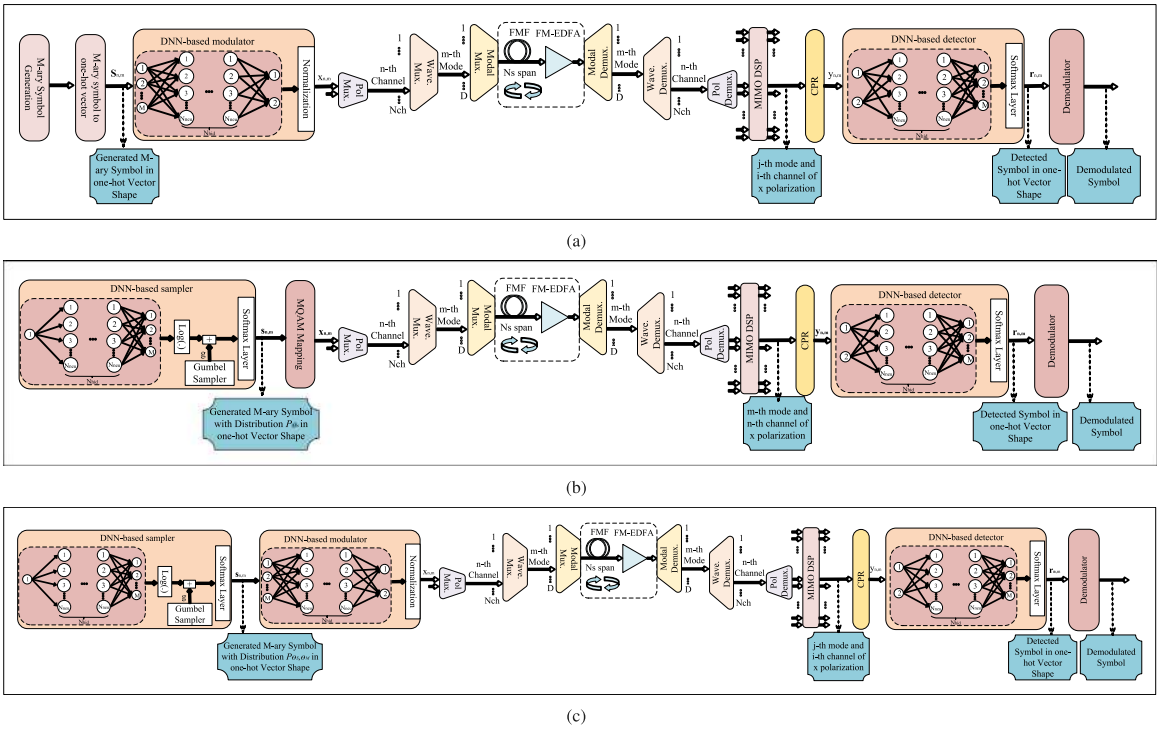

Fig. 1. The proposed EEDL-based (a) GCS algorithm, (b) PCS algorithm, and (c) JGPCS algorithm for FMF system.

3.2. PCS

The proposed EEDL-based PCS algorithm described by Fig. 1b is composed of two DNN-based functions representing sampler and detector. The DNN-based sampler must learn the parametric distribution $p_{\theta_S}(\boldsymbol{s_{n,m}})$ for transmitted symbols of nth channel and mth mode which provided to the DNN-based detector, holds enough information for replicating the transmitted symbols at the receiver output. The DNN-based sampler is composed of an input layer with 1 neuron, an output layer with M neurons, and N_{hid} hidden layers which each of them has N_{neu} hidden neurons. The DNN-based sampler tries to find the optimum probabilistic distribution in an end-to-end process based on the channel condition which is described by EGN-model and is in turn dependent on the input power. Therefore, we use the input power as the only input feature of DNN-based sampler. The training of PCS is done by optimizing an end-to-end loss function which is dependent on the mathematical model of all processing blocks, i.e., sampler, modulator, channel and detector. The SGD is used for minimizing this loss function, accordingly, the mathematical model of all these blocks should be differentiable. The mathematical models of modulator, detector, and channel are differentiable functions. Therefore, the challenge of employing EEDL-based algorithms for PCS is to design a trainable sampler which is differentiable w.r.t. parametric distribution $p_{\theta_S}(\boldsymbol{s_{n,m}})$ (i.e., the sampler should not have functions like max(.), |.| and so on). It is shown in [45] that a convenient way to sample a discrete distribution $p_{\theta_S}(\boldsymbol{s_{n,m}})$ is to find the maximizing argument of the summation of sample with Gumble distribution (g) and $log(p_{\theta_S}(\boldsymbol{s_{n,m}}))$. In other words, the mathematical model for computing the samples is

$$\boldsymbol{s_{n,m}} = \max_i(g_i + log(p_{\theta_S}(i))). \tag{13}$$

The max operator is not differentiable, thus, the SGD method cannot be used. Therefore, we consider Gumbel-Softmax trick which uses softmax as an approximate of max, and rewrite (13) as

$$\boldsymbol{s_{n,m}} \approx \max_i \frac{exp^{\frac{g_i+log(p_{\theta_S}(i))}{\tau}}}{\sum_{j=1}^{S} exp^{\frac{g_i+log(p_{\theta_S}(i))}{\tau}}}; i = 1, \ldots, N, \tag{14}$$

where τ is a positive parameter, if $\tau \to 0$ then $\boldsymbol{s_{n,m}}$ approaches to one-hot vector with a distribution close to $p_{\theta_S}(\boldsymbol{s_{n,m}})$ [45,46]. In this work, we take the same approach as [39–41] and assume that a distribution matcher exists which maps the bits to symbols according to the distribution $p_{\theta_S}(\boldsymbol{s_{n,m}})$. Note that in order to be practically applied to the actual FMF system, after obtaining the optimal distribution in DNN-based sampler, a distribution matcher e.g., conventional algorithms as comprehensively investigated in [47] can convert the uniform distribution bit sequence to the target distribution symbol sequence. Then $\boldsymbol{s_{n,m}}$ is mapped on $x_{n,m}$ based on the following formulation

$$x_{n,m} = f_{QAM}(\boldsymbol{s_{n,m}}), \tag{15}$$

where $f_{QAM}(.)$ is the QAM modulator. Therefore, $x_{n,m}$ has the following distribution

$$p_{\theta_S}(x_{n,m}) = \sum_{s=1}^{N} \delta(x_{n,m} - f_{QAM}(\boldsymbol{s_{n,m}}))p_{\theta_S}(\boldsymbol{s_{n,m}}), \tag{16}$$

where $\delta(.)$ denotes the Dirac delta. The modulated signal of is transmitted via FMF link as described by

$$y_{n,m} = f_{EGN}(x_{n,m}). \tag{17}$$

The detected symbol, $\boldsymbol{r_{n,m}}$, is obtained by passing $y_{n,m}$ through the DNN-based detector based on the following equation

$$\boldsymbol{r_{n,m}} = f_{\theta_D}(y_{n,m}). \tag{18}$$

The DNN-based detector, $f_{\theta_D}(.)$, has the same structure as described in Section 3.1, and should learn the mapping denoted by $\tilde{p}_{\theta_D}(s_{n,m}|y_{n,m})$ which is an approximate of the true posterior distribution $p_{\theta_S}(s_{n,m}|y_{n,m})$. Therefore, we define the following loss

function

$$L(\boldsymbol{\theta}) = E_{s_{n,m},y_{n,m}}\{-log(\tilde{p}_{\boldsymbol{\theta_D}}(s_{n,m}|y_{n,m}))\}$$
$$= -\sum_{s_{n,m}=1}^{M} p_{\boldsymbol{\theta_S}}(\boldsymbol{s_{n,m}}) \int_y p(y_{n,m}|f_{QAM}(\boldsymbol{s_{n,m}}))log(\tilde{p}_{\boldsymbol{\theta_D}}(s_{n,m}|y_{n,m}))dy =$$
$$-\int_x p_{\boldsymbol{\theta_S}}(x_{n,m}) \int_y p(y_{n,m}|x_{n,m})log(\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))dydx$$
$$\overset{(1)}{=} -\int_x \int_y p_{\boldsymbol{\theta_S}}(x_{n,m}, y_{n,m})log(\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))dydx \overset{(2)}{=}$$
$$-\int_x p_{\boldsymbol{\theta_S}}(x_{n,m})log(p_{\boldsymbol{\theta_S}}(x_{n,m}))dx - \int_x \int_y p_{\boldsymbol{\theta_S}}(x_{n,m}, y_{n,m})$$
$$\times\ log\left(\frac{\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}, y_{n,m})}{p_{\boldsymbol{\theta_S}}(y_{n,m})p_{\boldsymbol{\theta_S}}(x_{n,m})}\right)dydx, \tag{19}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta_S}, \boldsymbol{\theta_D}\}$, equality (1) is based on $p_{\boldsymbol{\theta_S}}(x_{n,m}, y_{n,m}) = p_{\boldsymbol{\theta_S}}(x_{n,m})p(y_{n,m}|x_{n,m})$, and equality (2) is based on $p_{\boldsymbol{\theta_S}}(y_{n,m}) = \int_x p_{\boldsymbol{\theta_S}}(x_{n,m}|y_{n,m})$ and $\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}, y_{n,m}) = \tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m})p_{\boldsymbol{\theta_S}}(y_{n,m})$ with $p_{\boldsymbol{\theta_S}}(x_{n,m}, y_{n,m})$ as true joint distribution of $(x_{n,m}, y_{n,m})$ and $\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}, y_{n,m})$ as joint distribution according to $\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m})$. The optimum $\boldsymbol{\theta}$ can be obtained by iteratively minimizing (19) based on (12) and using SGD method. (19) can be written as

$$L(\boldsymbol{\theta}) = H_{\boldsymbol{\theta_S}}(S_{n,m}) - I_{\boldsymbol{\theta_S}}(X_{n,m}; Y_{n,m}) + E_{y_{n,m}}\{D_{KL}(p_{\boldsymbol{\theta_S}}(x_{n,m}|y_{n,m}) \parallel$$
$$\times\ \tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))\}, \tag{20}$$

where $H_{\boldsymbol{\theta_S}}(\boldsymbol{S_{n,m}})$ is the entropy of $\boldsymbol{S_{n,m}} = \{s^1_{n,m}, \ldots, s^N_{n,m}\}$, $I_{\boldsymbol{\theta_S}}(X_{n,m}; Y_{n,m})$ is the MI between the channel input $X_{n,m} = \{x^1_{n,m}, \ldots, x^N_{n,m}\}$ and the channel output $Y_{n,m} = \{y^1_{n,m}, \ldots, y^N_{n,m}\}$, which constitutes the maximum achievable information rate under the FMF nonlinear channel [48], $D_{KL}(p_{\boldsymbol{\theta_S}}(x_{n,m}|y_{n,m}) \parallel \tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))$ is the Koulback–Liber (KL) divergence between the true posterior distribution, $p_{\boldsymbol{\theta_S}}(x_{n,m}|y_{n,m})$, and approximated posterior distribution, $\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m})$. Minimizing the loss function is in touch with joint maximizing MI and minimizing KL divergence. The DNN-based detector approximates the true posterior distribution that maximizes MI which in turn reduces the possibility of achieving a constellation that well approximates posterior distribution but do not maximizes MI. Thus, the DNN-based detector should learn to approximate a wide range of posterior distributions with high precision.

3.3. JGPCS

Proposed EEDL-based JGPCS algorithm described by Fig. 1c is composed of three DNN-based functions, a sampler, a modulator and a detector. The DNN-based sampler must learn the parametric distribution $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(\boldsymbol{s_{n,m}})$ for transmitted symbols of nth channel and mth mode which provided to the DNN-based modulator and DNN-based detector, holds enough information for replication the transmitted symbols at the receiver output. The sampler has the same structure as Section 3.2. Considering the Gumbel-Softmax trick, the symbol $\boldsymbol{s_{n,m}}$ with parametric distribution $p_{\boldsymbol{\theta_S}}(\boldsymbol{s_{n,m}})$ can be obtained by

$$\boldsymbol{s_{n,m}} \approx \max_i \frac{exp(g_i + log(p_{\boldsymbol{\theta_S}}(i))/\tau)}{\sum_{j=1}^{S} exp(g_i + log(p_{\boldsymbol{\theta_S}}(i))/\tau)}; i = 1, \ldots, N. \tag{21}$$

Then $\boldsymbol{s_{n,m}}$ is modulated to constellation point $x_{n,m}$ via the following function

$$x_{n,m} = f_{\boldsymbol{\theta_M}}(\boldsymbol{s_{n,m}}). \tag{22}$$

The DNN-based modulator, $f_{\boldsymbol{\theta_M}}(.)$, has the same structure as Section 3.1. The modulated signal of is transmitted via FMF link as described by

$$y_{n,m} = f_{EGN}(x_{n,m}). \tag{23}$$

The detected symbol, $\boldsymbol{r_{n,m}}$, is obtained by passing $y_{n,m}$ through the DNN-based detector based on the following equation

$$\boldsymbol{r_{n,m}} = f_{\boldsymbol{\theta_D}}(y_{n,m}). \tag{24}$$

The DNN-based detector, $f_{\boldsymbol{\theta_D}}(.)$, has the same structure as described in Section 3.1, and should learn the mapping denoted by $\tilde{p}_{\boldsymbol{\theta_D}}(s_{n,m}|y_{n,m})$ which is the approximation of the true posterior distribution $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(s_{n,m}|y_{n,m})$. We define the following loss function

$$L(\boldsymbol{\theta}) = E_{s_{n,m},y_{n,m}}\{-log(\tilde{p}_{\boldsymbol{\theta_D}}(s_{n,m}|y_{n,m}))\} = -\sum_{s_{n,m}=1}^{M} p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(s_{n,m})$$
$$\times \int_y p(y_{n,m}|f_{\boldsymbol{\theta_M}}(s_{n,m}))log(\tilde{p}_{\boldsymbol{\theta_D}}(s_{n,m}|y_{n,m}))dy =$$
$$-\int_x p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}) \int_y p(y_{n,m}|x_{n,m})log(\ p_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))dydx$$
$$\overset{(1)}{=} -\int_x \int_y p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}, y_{n,m})log(\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))dydx \overset{(2)}{=}$$
$$-\int_x p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m})log(p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}))dx - \int_x \int_y p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}, y_{n,m})$$
$$\times\ log(\frac{\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}, y_{n,m})}{p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(y_{n,m})p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m})})dydx, \tag{25}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta_S}, \boldsymbol{\theta_M}, \boldsymbol{\theta_D}\}$, equality (1) is based on $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}, y_{n,m}) = p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m})p(y_{n,m}|x_{n,m})$, and equality (2) is based on $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(y_{n,m}) = \int_x p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}|y_{n,m})$ and $\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}|y_{n,m}) = \tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m})p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(y_{n,m})$ with $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}, y_{n,m})$ as true joint distribution of $(x_{n,m}, y_{n,m})$ whereas $\tilde{p}_{\boldsymbol{\theta}}(x_{n,m}, y_{n,m})$ is the joint distribution computed from the posterior approximated by the detector $\tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m})$. The optimum $\boldsymbol{\theta}$ can be obtained by iteratively minimizing (26) based on (12) and using SGD method. (26) can be written as

$$L(\boldsymbol{\theta}) = H_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(\boldsymbol{S_{n,m}}) - I_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(X_{n,m}; Y_{n,m})$$
$$+ E_{y_{n,m}}\{D_{KL}(p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}|y_{n,m}) \parallel \tilde{p}_{\boldsymbol{\theta_D}}(x_{n,m}|y_{n,m}))\} \tag{26}$$

where $p_{\boldsymbol{\theta_S},\boldsymbol{\theta_M}}(x_{n,m}|y_{n,m})$ is the true posterior distribution of $x_{n,m}$, the same insights as (20) goes here.

4. Simulation results

This section presents and compares the simulation results of the proposed EEDL-based GCS and PCS, JGPCS, and QAM constellation. In simulations we consider a FMF optical link with 3 spatial modes and 5 WDM channels with the center frequency 1550 nm. The transmitted symbol rate in each WDM channel is 32 GBaud with 50 GHz channel spacing. 10 spans each with 100 km length are deployed. Moreover, the considered modulation order is 64. The nonlinear coupling coefficient, modal dispersion, chromatic dispersion, and attenuation values are shown in Tables 1 and 2. FM-EDFA noise figure is 5 dB. We deployed the simulations in Python/Tensorflow and tune the hyperparameters manually based on [49–51]. The tuned hyperparameters are represented in Table 3.

Fig. 2 plots the MI versus launched power per channel-mode, for proposed EEDL-based GCS, PCS, and JGPCS algorithms, uniform distributed QAM and Maxwell–Boltzmann distributed QAM, considering central channel of LP01 mode. The proposed EEDL-based

Table 1
Nonlinear coupling coefficient (γf_{mq} (1/W/km)) [4].

mq	LP01	LP11a	LP11b
LP01	0.73	0.36	0.36
LP11a	0.36	0.18	0.55
LP11b	0.36	0.18	0.18

Table 2
Attenuation ($\alpha_{n,m}$ (dB/km)), and dispersion terms (β_{1_m} (ps/km), β_{2_m} (ps^2/km), and β_{3_m} (ps^3/km)) [4].

	LP01	LP11a	LP11b
$\alpha_{n,m}$	0.2	0.2	0.2
β_{1_m}	−0.29	−0.66	−0.66
β_{2_m}	28.27	26.96	26.96

Table 3
The tuned hyperparameters used in simulations of this paper.

Hyperparameter	Symbol	Value
Number of hidden layers	N_{hid}	2
Number of hidden neurons	N_{neu}	32
Batch size	K	2^{16}
Sample size to batch size ratio	η_s	4
Number of iterations	N_{iter}	700
Activation function	$\alpha(.)$	Relu/Softmax
Loss function	$L(.)$	Cross-entropy
Learning rate	η	0.001
Optimizer	–	Adam

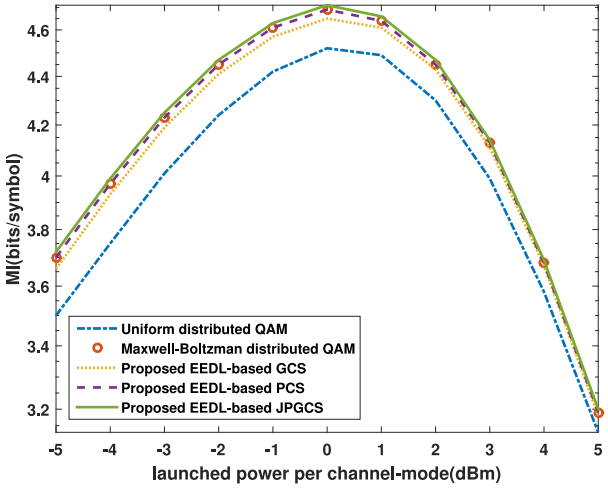

Fig. 2. MI versus launched power per channel-mode, for proposed EEDL-based GCS, PCS, and JGPCS algorithms, uniform distributed QAM and Maxwell–Boltzmann distributed QAM, considering central channel of LP01 mode.

algorithms mitigate both ASE and NLI noise effect by optimizing the location and/or occurrence probability of constellation points. The learned constellations show improved MI performance compared with uniform distributed QAM constellation over all power regimes. Considering the optimum power, this improvement (the shaping gain) is 0.15, 0.19, and 0.22 bits/symbol respectively for proposed EEDL-based GCS, PCS, and JGPCS algorithms. We observe the same shaping gain at different powers in low power regime where ASE noise is dominant. However, at high power regime where NLI noise is dominant the shaping gain is heavily reduced and vanishes for highly nonlinear region. The proposed EEDL-based constellation shaping has the same MI as the well-known Maxwell–Boltzmann distributed QAM constellation which is optimum in an AWGN channel. However, the Maxwell–Boltzmann distributed QAM, although optimal, has

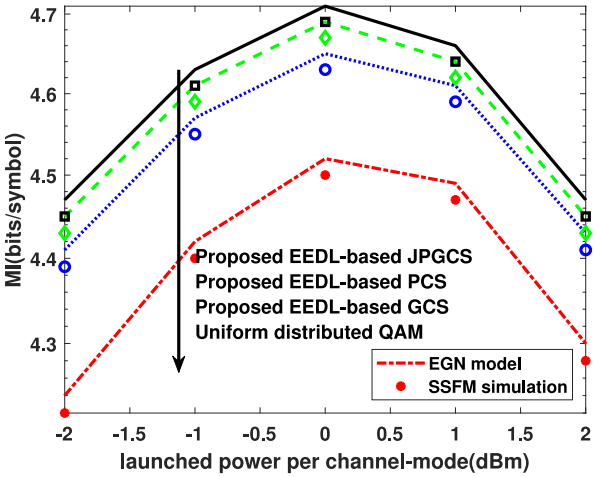

Fig. 3. MI versus launched power per channel-mode, for proposed EEDL-based GCS, PCS, and JGPCS algorithms, uniform distributed QAM based on EGN model and SSFM simulation, considering central channel of LP01 mode.

some barriers while practical implementation. For instance, the ratio between the probability of occurrence of the inner and the outer constellation points of the Maxwell–Boltzmann distributed QAM constellation is almost 10^5 while in EEDL-based PCS it is around 10^3. This means that the proposed EEDL-based PCS is more practical to precisely produce a sequence with an occurrence probability that fits in a DAC's memory. MI improvement in JPCS is 15% more than MI improvement in PCS. The PCS and JGPCS have the same complexity as the practical implementation of both QAM modulator and DNN-based modulator is a look-up table. Therefore, proposing JGPCS becomes meaningful while considering JGPCS obtains 15% more MI improvement than PCS, without any additional complexity, latency, or cost.

The well-known split-step Furrier method (SSFM) approximates the FMF outcome by solving the Manakov equation through many consecutive numerical simulation steps. The SSFM is verified by some experiments [52] reflecting the real transmission environment. Therefore, we employ an SSFM simulation in Fig. 3 based on instructions provided by [6,17] to verify that the obtained results based on EGN model truly reflect the performance improvement of the proposed algorithms. As seen, close agreement can be observed with a very small gap which is due to the fact that for calculating MI for EGN model and SSFM simulation we use (A.2) and (A.4), respectively. More explanations about this observation is provided in [16,30] and in Appendix. However, the important thing to note is that the same MI improvements as in EGN model are provided by SSFM simulation results.

Fig. 4 depicts the learned constellations for proposed EEDL-based GCS (top), PCS (center), and JGPCS (bottom) algorithms, for −10 dBm (left), optimum (middle), and 10 dBm (right) launched power per channel-mode, considering central channel of LP01 mode. In GCS, at low power regime, the inner constellation points are randomly positioned after training. The NLI noise is the FMF dominant effect at high launched power regime wherein the learned constellation points locate either at the origin or on a uniform ring. In PCS, the color of the constellation points represents their occurrence probabilities. The yellow points are with the highest and the blue point are the lowest occurrence probabilities. At low power regime, the learned distributions look like a circular complex Gaussian distribution. The symbols with less occurrence probabilities are located further apart from the origin which is due to the normalization constraint ($E|x|^2 = 1$). At high power regime, the learned distribution is similar to a

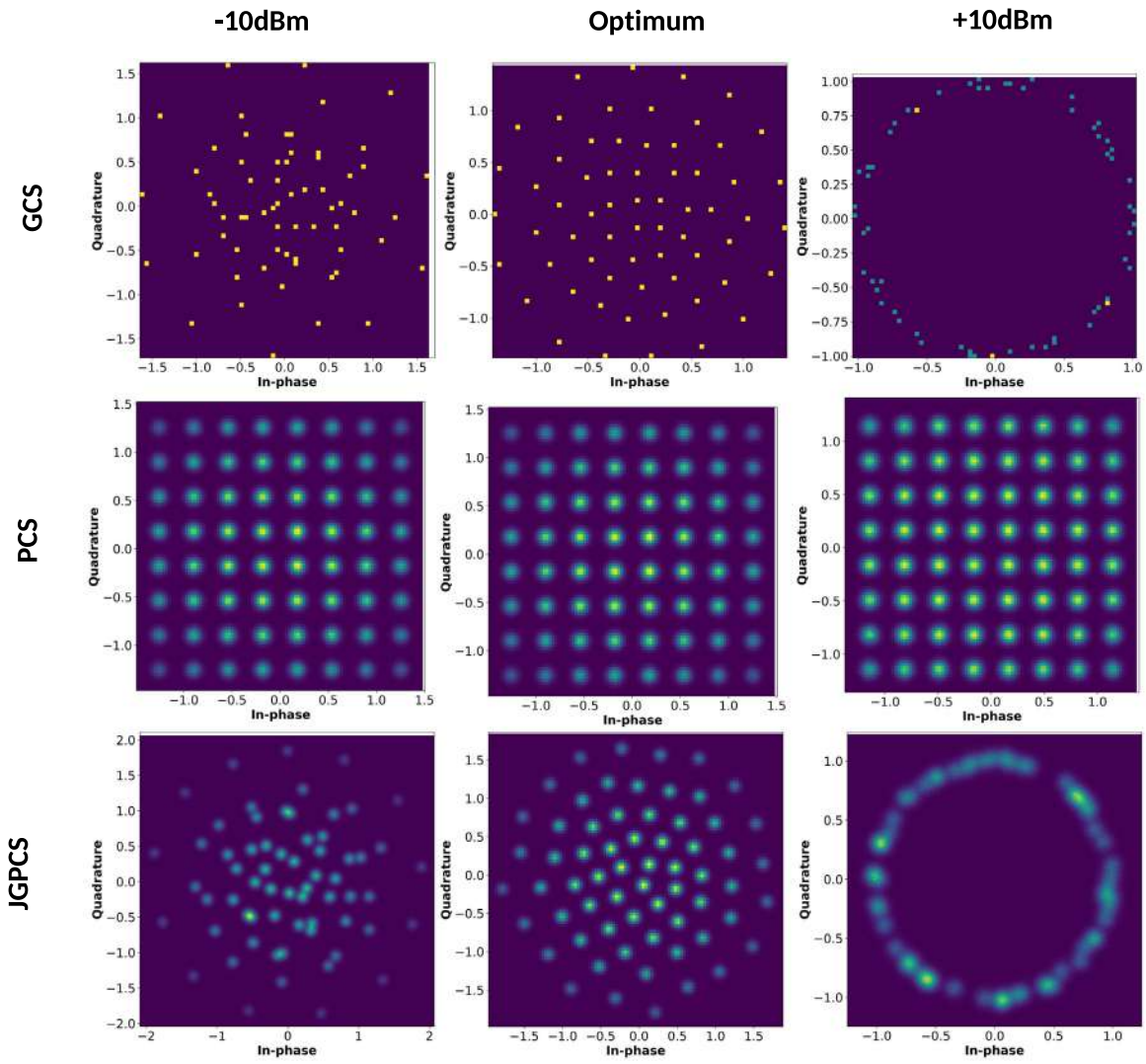

Fig. 4. Learned constellations for proposed EEDL-based GCS (top), PCS (center), and JGPCS (bottom) algorithms, for −10 dBm (left), optimum (middle), and 10 dBm (right) launched power per channel-mode, considering central channel of LP01 mode.

uniform distribution. The inner learned constellation points have lower-power and are more frequently used as opposite to the outer points with higher power which in turn enable the learned constellation to be tolerated to NLI noise. In JGPCS, again, we see that the constellation points are distributed inside a Gaussian-like constellation. At low and high launched power per channel-mode regimes, constellation points are closer to the origin. The same insights as the learned GCS and PCS constellations goes here.

5. Discussion

As seen in Fig. 1(b), we only feed the DNN-based sampler a one-dimensional input, however, results indicate that the EEDL-based PCS algorithm can learn the optimum distribution properly. In the following, we discuss about this from different aspects. The main reason that EEDL-based PCS can learn the optimum distribution properly even with a one-dimensional input is due to training by backward propagation. In simulations, we first define the launched power and feed into DNN-based sampler, after constellation mapping, the transmitter output is propagated through FMF channel, and entered DNN-based detector. We formulate the loss function inside which the whole propagation procedures are inherent. The backward propagation computes the loss function gradient with respect to each weight, calculates gradient one layer at a time, and iterates backward from last layer. In other words, everything (the effect of FMF channel, as well as the impact of transmitter and receiver processing blocks) is being considered while adjusting the DNN-based sampler weights and not only the one-dimensional input. The EEDL-based PCS idea is to find the optimum probability distribution for a pre-specified launched power. The launched power (in dBm) is fed to a DNN with trainable parameters $\boldsymbol{\theta_S}$, and DNN-based sampler provides a continuum of distributions $p_{\boldsymbol{\theta_S}}$ that are determined by the launched power. Available works on EEDL-based PCS feed the DNN-based sampler a one-dimensional input, (SNR in dB) (see [39–41]). The motivation is that the optimum probabilistic distribution depends on the SNR of the respective channel (see [53]). Note that these works consider radio frequency channel and feedback the SNR from receiver to transmitter while feedback of SNR is not practical in long-range FMF links. Since the channel coherency of FMF (due to linear coupling) is much

less than radio frequency channel and link-range of FMF is much more than radio frequency channel which in turn results in outdated feedback information [54]. SNR in FMF link is in turn dependent on the launched power, span length, number of spans, attenuation, dispersion, and nonlinearity coefficient [4]. Here, all these parameters except launched power are fixed as we consider a point-to-point FMF link, and feeding fixed parameters into DNN only increases the complexity and does not improve performance. Launched power is a variable known by transmitter and thus we feed it to the DNN-based sampler.

6. Conclusion

FMF NLI noise is the main practical limitation in front of long-haul FMF optical communication. Providing an optimal constellation by designing the location and occurrence probability of constellation points can reduce the performance degradation due to FMF nonlinearity. To this aim, in this paper, we have proposed EEDL-based GCS, PCS, and JGPCS algorithms for FMF system. We have compared the proposed EEDL-based algorithms with uniform distributed QAM constellation and the well-known Maxwell–Boltzmann distributed QAM constellation. Simulation results have shown proposed EEDL-based GCS, PCS, and JGPCS algorithms achieve MI improvement of 0.15, 0.19, and 0.22 bits/symbol, respectively, compared with uniform distributed QAM constellation.

CRediT authorship contribution statement

Mohammad Ali Amirabadi: Conception and design of study, Writing – original draft, Writing – review & editing. **Mohammad Hossein Kahaei:** Conception and design of study, Writing – original draft, Writing – review & editing. **S. Alireza Nezamalhosseini:** Conception and design of study, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

All authors approved the version of the manuscript to be published.

Appendix. MI estimation

We follow the same approach as [48,55] to estimate the symbol wise MI. Consider the channel input $X_{n,m}$ be 2D random variable drawn from constellation $X_{n,m} = \{x^1_{n,m}, \ldots, x^N_{n,m}\}$, and the channel output $Y_{n,m}$ be 2D random variable. The MI between $X_{n,m}$ and $Y_{n,m}$ can be calculated by [48,55]

$$I(X_{n,m}; Y_{n,m}) = E_{X_{n,m}Y_{n,m}}[log_2(\frac{p_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})}{p_{Y_{n,m}}(Y_{n,m})})], \tag{A.1}$$

where $p_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ is the conditional channel pdf and $p_{Y_{n,m}}(Y_{n,m})$ is the channel output pdf, and MI is defined in units of bit per 2D symbol (bits/symbol). In reality, FMF channel exhibits memory wherein (A.1) becomes a lower bound on the MI, however, using the MIMO DSP at the receiver compensates FMF memory, therefore, we focus on memoryless MI. (A.1) can be solved by Monte-Carlo integration and based on the weak law of large numbers we have

$$\frac{1}{N}\sum_{i=1}^{N} log_2(\frac{p_{Y_{n,m}|X_{n,m}}(y^i_{n,m}|x^i_{n,m})}{p_{Y_{n,m}}(y^i_{n,m})}) \xrightarrow{p} I(X_{n,m}; Y_{n,m}), \tag{A.2}$$

where N is number of input–output pairs, and $\xrightarrow{(p)}$ denotes convergence in probability. Note that the channel pdf is known while dealing with the EGN-model and (A.2) estimates the MI whose accuracy increases with N. However, the channel pdf is not known while dealing with SSFM, and (A.1) cannot be evaluated directly as (A.2). As a solution, a lower bound on MI is achieved in [48,55] by considering an auxiliary channel pdf $q_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ instead of the true yet unknown channel pdf $p_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ which yields to

$$\begin{aligned} I(X_{n,m}; Y_{n,m}) &= E_{X_{n,m}Y_{n,m}}[log_2(\frac{p_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})}{p_{Y_{n,m}}(Y_{n,m})})] \\ &\geqslant E_{X_{n,m}Y_{n,m}}[log_2(\frac{q_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})}{q_{Y_{n,m}}(Y_{n,m})})] \triangleq R. \end{aligned} \tag{A.3}$$

(A.3) can be solved by Monte-Carlo integration and based on the weak law of large numbers we have

$$\frac{1}{N}\sum_{i=1}^{N} log_2(\frac{q_{Y_{n,m}|X_{n,m}}(y^i_{n,m}|x^i_{n,m})}{q_{Y_{n,m}}(y^i_{n,m})}) \xrightarrow{p} R. \tag{A.4}$$

MI and its lower bound are both the achievable rates, the better $q_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ resembles $p_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ the tighter this bound and the higher achievable rate. $q_{Y_{n,m}|X_{n,m}}(Y_{n,m}|X_{n,m})$ is 2D Gaussian distributed, comprehensive details about this selection is provided by [48,55]. (A.4) is not the true MI, however, this approach provides a practical achievable rate since a decoder should also assume a channel.

References

[1] R.J. Essiambre, Nonlinear capacity limit to optical communications, in: Nonlinear Optics, 2015, pp. NTu2A–3.
[2] Y. Fazea, M.S. Sajat, A. Ahmad, M.M. Alobaedy, Channel optimization in mode division multiplexing using neural networks, in: 14th International Colloquium on Signal Processing and Its Applications, 2018, pp. 173–175.
[3] A. Juarez, E. Krune, S. Warm, C. Bunge, K. Petermann, Modeling of mode coupling in multimode fibers with respect to bandwidth and loss, J. Lightwave Technol. 32 (8) (2014) 1549–1558.
[4] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Lawrence R. Chen, Joint power and gain allocation in MDM-WDM optical communication networks based on enhanced Gaussian noise model, IEEE Access 10 (2022) 23122–23139.
[5] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, F. Arepanaei, A. Carena, Closed-form EGN model for FMF systems, in: ACP, 2021.
[6] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Lawrence R. Chen, Optimal power allocation in nonlinear MDM-WDM systems using Gaussian noise model, IET Optoelectron. (2022).
[7] R.T. Jones, M.P. Yankov, D. Zibar, End-toend learning for GMI optimized geometric constellation shape, 2019, arXiv preprint arXiv:1907.08535.
[8] R.-J. Essiambre, G. Kramer, P.J. Winzer, G.J. Foschini, B. Goebel, Capacity limits of optical fiber networks, J. Lightw. Technol. 28 (4) (2010) 662–701.
[9] I.B. Djordjevic, H.G. Batshon, L. Xu, T. Wang, Coded polarizationmultiplexed iterative polar modulation (PM-IPM) for beyond 400 Gb/s serial optical transmission, in: Proc. Opt. Fiber. Conf, San Diego, CA, USA, 2010.
[10] T.H. Lotz, X. Liu, S. Chandrasekhar, P.J. Winzer, H. Haunstein, S. Randel, S. Corteselli, B. Zhu, D.W. Peckham, Coded PDM-OFDM transmission with shaped 256- iterative-polar-modulation achieving 11.15-b/s/Hz intrachannel spectral efficiency and 800-km reach, J. Lightwave Technol. 31 (4) (2013) 538–545.
[11] J. Cho, P.J. Winzer, Probabilistic constellation shaping for optical fiber communications, J. Lightwave Technol. 37 (6) (2019) 1590–1607.
[12] F.R. Kschischang, S. Pasupathy, Optimal nonuniform signaling for gaussian channels, IEEE Trans. Inform. Theory 39 (3) (1993) 913–929.

[13] D. Pilori, G. Bosco, C. Fludger, Impact of finite-resolution DAC and ADC on probabilistically-shaped QAM constellations, in: 2017 IEEE Photonics Conference, IPC, IEEE, 2017, pp. 433–434.
[14] D. Pilori, L. Bertignono, A. Nespola, F. Forghieri, G. Bosco, Comparison of probabilistically shaped 64QAM with lower cardinality uniform constellations in long-haul optical systems, J. Lightwave Technol. 36 (2) (2018) 501–509.
[15] J.-X. Cai, et al., 70.46 Tb/s over 7, 600 km and 71.65 Tb/s over 6, 970 km transmission in C+L band using coded modulation with hybrid constellation shaping and nonlinearity compensation, J. Lightwave Technol. 36 (1) (2018) 114–121.
[16] R.T. Jones, T.A. Eriksson, Y.P. Metodi, D. Zibar, Deep learning of geometric constellation shaping including fiber nonlinearities, in: Proc. Eur. Conf. Opt. Commun, Rome, Italy, 2018.
[17] T. Koike-Akino, Y. Wang, D.S. Millar, K. Kojima, K. Parsons, Neural turbo equalization: Deep learning for fiber-optic nonlinearity compensation, J. Lightwave Technol. 38 (11) (2020) 3059–3066.
[18] J. Du, T. Yang, X. Chen, J. Chai, Y. Zhao, S. Shi, A CNN-based cost-effective modulation format identification scheme by low-bandwidth direct detecting and low rate sampling for elastic optical networks, Opt. Commun. 471 (2020) 126007.
[19] H. Lv, X. Zhou, J. Huo, J. Yuan, Joint OSNR monitoring and modulation format identification on signal amplitude histograms using convolutional neural network, Opt. Fiber Technol., Mater. Devices Syst. 61 (2021) 102455.
[20] B. Karanov, M. Chagnon, F. Thouin, T.A. Eriksson, H. Bulow, D. Lavery, P. Bayvel, L. Schmalen, Endto-end deep learning of optical fiber communications, J. Lightwave Technol. 36 (20) (2018) 4843–4855.
[21] Z. Yang, W. Yu, G. Peng, Y. Liu, L. Zhang, Recent progress on novel DSP techniques for mode division multiplexing systems: A review, Appl. Sci. 11 (4) (2021) 1363.
[22] B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, C. Moser, Multimode optical fiber transmission with a deep learning network, Light: Sci. Appl. 7 (1) (2018) 1–11.
[23] Y. An, L. Huang, J. Li, J. Leng, L. Yang, P. Zhou, Learning to decompose the modes in few-mode fibers with deep convolutional neural network, Opt. Express 27 (7) (2019) 10127–10137.
[24] P. Fan, M. Ruddlesden, Y. Wang, L. Zhao, C. Lu, L. Su, Learning enabled dense space-division multiplexing through a single multimode fibre, 2020, arXiv preprint arXiv:2002.01788.
[25] B. Poudel, J. Oshima, H. Kobayashi, K. Iwashita, MIMO detection using a deep learning neural network in a mode division multiplexing optical transmission system, Opt. Commun. 440 (2019) 41–48.
[26] V. Talreja, T. Koike-Akino, Y. Wang, D.S. Millar, K. Kojima, K. Parsons, End-to-end deep learning for phase noise-robust multi-dimensional geometric shaping, in: European Conference on Optical Communications, 2020, pp. 1–3.
[27] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, Properties of nonlinear noise in long, dispersion-uncompensated fiber links, Opt. Express 21 (22) (2013) 25685–25699.
[28] O. Geller, R. Dar, M. Feder, M. Shtaif, A shaping algorithm for mitigating inter-channel nonlinear phase-noise in nonlinear fiber systems, J. Lightwave Technol. 34 (16) (2016) 3884–3889.
[29] R. Dar, M. Feder, A. Mecozzi, M. Shtaif, Accumulation of nonlinear interference noise in fiber-optic systems, Opt. Express 22 (12) (2014) 14199–14211.
[30] R.T. Jones, T.A. Eriksson, M.P. Yankov, B.J. Puttnam, G. Rademacher, R.S. Luis, D. Zibar, Geometric constellation shaping for fiber optic communication systems via end-to-end learning, 2018, arXiv preprint arXiv:1810.00774.
[31] E. Sillekens, D. Semrau, D. Lavery, P. Bayvel, R.I. Killey, Experimental demonstration of geometrically-shaped constellations tailored to the nonlinear fibre channel, in: European Conference on Optical Communications, 2018, p. Tu3G.3.
[32] F. Steiner, G. Böcherer, Comparison of geometric and probabilistic shaping with application to ATSC 3.0, in: SCC 2017; 11th International ITG Conference on Systems, Communications and Coding, VDE, 2017, pp. 1–6.
[33] S. Zhang, F. Yaman, E. Mateo, T. Inoue, K. Nakamura, Y. Inada, Design and performance evaluation of a GMI-optimized 32qam, in: 2017 European Conference on Optical Communication, ECOC, IEEE, 2017, pp. 1–3.
[34] B. Chen, C. Okonkwo, H. Hafermann, A. Alvarado, Increasing achievable information rates via geometric shaping, in: 2018 European Conference on Optical Communication, ECOC, IEEE, 2018, pp. 1–3.
[35] O. Jovanovic, M.P. Yankov, F. Da Ros, D. Zibar, End-to-end learning of a constellation shape robust to variations in SNR and laser linewidth, in: 2021 European Conference on Optical Communication, ECOC, IEEE, 2021, pp. 1–4.
[36] S. Gaiarin, F. Da Ros, R.T. Jones, D. Zibar, End-to-end optimization of coherent optical communications over the split-step Fourier method guided by the nonlinear Fourier transform theory, J. Lightwave Technol. 39 (2) (2020) 418–428.
[37] S. Li, C. Häger, N. Garcia, H. Wymeersch, Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning, in: 2018 European Conference on Optical Communication, ECOC, IEEE, 2018, pp. 1–3.
[38] K. Gümüş, A. Alvarado, B. Chen, C. Häger, E. Agrell, End-to-end learning of geometrical shaping maximizing generalized mutual information, in: 2020 Optical Fiber Communications Conference and Exhibition, OFC, IEEE, 2020, pp. 1–3.
[39] V. Aref, M. Chagnon, End-to-end learning of joint geometric and probabilistic constellation shaping, in: 2022 Optical Fiber Communications Conference and Exhibition, OFC, IEEE, 2018, pp. 1–3.
[40] M. Stark, F.A. Aoudia, J. Hoydis, Joint learning of geometric and probabilistic constellation shaping, in: 2019 IEEE Globecom Workshops (GC Wkshps), IEEE, 2019, pp. 1–6.
[41] F.A. Aoudia, J. Hoydis, Joint learning of probabilistic and geometric shaping for coded modulation systems, in: GLOBECOM 2020-2020 IEEE Global Communications Conference, IEEE, 2020, pp. 1–6.
[42] S. Mumtaz, R.J. Essiambre, G.P. Agrawal, Nonlinear propagation in multimode and multicore fibers: generalization of the Manakov equations, J. Lightwave Technol. 31 (3) (2012) 398–406.
[43] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, A deep learning based detector for FSO system considering imperfect CSI scenario, in: 2020 3rd West Asian Symposium on Optical and Millimeter-Wave Wireless Communication, 2020, pp. 1–5.
[44] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Deep learning based detection technique for FSO communication systems, Phys. Commun. 43 (2020) 101229.
[45] E. Jang, S. Gu, B. Poole, Categorical reparameterization with Gumbel-Softmax, in: Proc. Int. Conf. Learn. Represent, 2017.
[46] Y. Bengio, N. Léonard, A. Courville, Estimating or propagating gradients through stochastic neurons for conditional computation, 2013, arXiv preprint arXiv:1308.3432.
[47] Y.C. Gültekin, T. Fehenberger, A. Alvarado, F.M. Willems, Probabilistic shaping for finite blocklengths: Distribution matching and sphere shaping, Entropy 22 (5) (2020) 581.
[48] T. Fehenberger, A. Alvarado, G. Bocherer, N. Hanik, On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel, J. Lightwave Technol. 34 (21) (2016) 5063–5073.
[49] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Novel suboptimal approaches for hyperparameter tuning of deep neural network [under the shelf of optical communication], Phys. Commun. 41 (2020) 101057.
[50] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, V.T. Vakili, Deep learning for channel estimation in FSO communication system, Opt. Commun. 459 (2020) 124989.
[51] M.A. Amirabadi, M.H. Kahaei, S.A. Nezamalhosseini, Low complexity deep learning algorithms for compensating atmospheric turbulence in the free space optical communication system, IET Optoelectron. (2021).
[52] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, F. Forghieri, The GN-model of fiber non-linear propagation and its applications, J. Lightwave Technol. 32 (4) (2013) 694–721.
[53] G. Böcherer, F. Steiner, P. Schulte, Bandwidth efficient and rate-matched low-density parity-check coded modulation, IEEE Trans. Commun. 63 (12) (2015) 4651–4665.
[54] S.O. Arik, J.M. Kahn, K.P. Ho, MIMO signal processing for mode-division multiplexing: An overview of channel models and signal processing architectures, IEEE Signal Process. Mag. 31 (2) (2014) 25–34.
[55] T.A. Eriksson, T. Fehenberger, P.A. Andrekson, M. Karlsson, N. Hanik, E. Agrell, Impact of 4D channel distribution on the achievable rates in coherent optical communication experiments, J. Lightwave Technol. 34 (9) (2016) 2256–2266.

Mohammad Ali Amirabadi was born in Zahedan, Iran, in 1993. He received the B.Sc. degree in Optics & Laser Engineering from Malek-e-Ashtar University of Technology, Isfahan, Iran, in 2015, and the M.Sc. degree in Communication Engineering from Iran University of Science and Technology, Tehran, Iran in 2017. Now he is studying Ph.D. in Communication Engineering in Iran University of Science and Technology, Tehran, Iran. His research interests include Multimode Fiber Optic Communication, Free Space Optical Communication, and Deep Learning.

Mohammad Hossein Kahaei received the B.Sc. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 1986, the M.Sc. degree in adaptive signal processing from the University of the Ryukyus, Okinawa, Japan, in 1994, and the Ph.D. degree in signal processing from Queensland University of Technology, Brisbane, Australia, in 1998. Since 1999, he has been with the School of Electrical Engineering, Iran University of Science and Technology, Tehran, Iran, where he is currently an Associate Professor and the Head of Signal and System Modeling laboratory. His research interests include array signal processing with primary emphasis on compressed sensing, sparse optimization problems, data science, localization, tracking, DOA estimation, blind source separation, and wireless sensor networks.

S. Alireza Nezamalhosseini received the B.Sc. degree in electrical engineering from Amirkabir University of Technology, Tehran, Iran, in 2006, and the M.Sc. and Ph.D. degrees in electrical engineering from Sharif University of Technology (SUT), Tehran, Iran, in 2008 and 2013, respectively. He is currently an assistant professor at Iran University of Science and Technology (IUST), Tehran. His research interests include underwater wireless optical communications, mode-division multiplexing in optical fibers, and visible light communications.