Geometric Constellation Shaping for Concatenated Two-Level Multi-Level Codes

Matsumine, Toshiki; Yankov, Metodi Plamenov; Forchhammer, Søren

Published in:
Journal of Lightwave Technology

Link to article, DOI:
10.1109/JLT.2022.3179529

Publication date:
2022

Document Version
Peer reviewed version

Geometric Constellation Shaping for Concatenated Two-Level Multi-Level Codes

Toshiki Matsumine, *Member, IEEE*, Metodi P. Yankov, *Member, IEEE*, and Søren Forchhammer, *Member, IEEE*.

***Abstract*—This paper studies a learning approach to geometric shaping (GS) for concatenated two-level multi-level codes (MLC). Unlike standard bit-interleaved coded modulation (BICM), the proposed MLC protects only the least reliable bits (LRBs) by an inner code, and thereby efficiently reduces complexity associated with inner soft-decision (SD) decoding compared to the BICM without performance degradation. We propose a new loss function for training a geometric constellation shape in the proposed MLC, leading to maximization of the achievable rate. More specifically, considering the different coding structures for the most reliable bits (MRBs) and the LRBs, the proposed loss function combines two different functions: the union bound on a bit error rate (BER) for the MRBs and the generalized mutual information (GMI) for the LRBs. We demonstrate by simulations of wavelength division multiplexing (WDM) optical fiber systems that the proposed loss function offers a performance gain over the conventional cross entropy-based loss function. Furthermore, it is demonstrated that the proposed MLC with GS outperforms the conventional BICM with GS in terms of a transmission distance, even with significantly lower SD decoding complexity.**

***Index Terms*—Multi-level codes, geometric shaping, neural network.**

I. Introduction

The use of powerful forward error correction (FEC) as well as high order modulation is inevitable for increasing optical network capacity. Specifically, a serial concatenation of hard-decision (HD) and soft-decision (SD) FECs has been adopted in modern optical communication systems such as 400ZR [1]. When concatenated with an outer HD decoder, the role of the inner SD decoder is simply to reduce the BER so the outer decoder can achieve a BER requirement specified by systems. It has been demonstrated in [2]–[6] that, for this role, two-level multi level codes (MLCs) [7], [8] that encode only the least reliable bit (LRB) of a mapper input by an inner FEC code achieve better performance-complexity trade-offs than the conventional bit-interleaved coded modulation (BICM) [9], [10] that protects all the bit levels equally by an inner code. Recently, similar two-level coded modulation approach has been also proposed in [11], [12] for four-dimensional lattices, termed Hurwitz constellations.

Constellation shaping is a technique to enhance the achievable rate by manipulating the signal distribution. There are largely two approaches to constellation shaping: probabilistic shaping (PS) and geometric shaping (GS). Contrary to PS that changes distribution of constellation points, GS changes the locations of constellation points without changing their distribution. Both PS and GS have been extensively studied for optical fiber communications [13]–[23], where the potential of constellation shaping for enhancing a transmission distance has been demonstrated.

In [24]–[27], applications of the probabilistic amplitude shaping (PAS) [28] to concatenated two-level MLC have been studied, where better performance-complexity trade-offs than the conventional BICM were demonstrated even with much lower complexity. Although PAS may achieve better achievable information rate (AIR) than GS in general [29], PAS requires an external distribution matcher (DM) whose efficient hardware implementation can be challenging especially for achieving a large shaping gain. Also, since the PAS was originally designed for Gray labeling, which is not necessarily well-suited for MLC, it is not straightforward to apply PAS to MLC [24]–[26]. In contrast, GS does not require a DM, and is compatible with two-level MLC.

Meanwhile, neural network-based design of communication signals, both PS and GS, has been extensively studied for various communication channels [17], [30]–[45]. In particular, for optical fiber communications, an end-to-end learning approach of communication systems, referred to as *autoencoder*, has been studied [34], [35], [37], [40], [41], [43]–[45]. However, the main focus of these existing works has been on BICM with SD FEC where a primary objective is to minimize the binary cross entropy loss between mapper input and demapper output.

In this paper, we propose a neural network-based design of GS for a concatenated FEC scheme with inner two-level MLC. Note that there exist several works that apply power-efficient lattice constellations to MLC, e.g., [11], [12], [46], which may be seen as GS. However, to the best of our knowledge, this is the first attempt to optimize GS for MLC by means of an autoencoder, which takes fiber nonlinearities into account unlike the conventional works. Furthermore, we propose a novel loss function for maximizing an AIR of the proposed MLC system. Specifically, assuming the nonlinear interference noise (NLIN) model for optical fiber communications [47], we train a symbol mapper for maximizing the generalized mutual information (GMI) of the LRBs and minimizing the BER of the MRBs. Since the BER is not differentiable, which is a necessary condition for a loss function in stochastic gradient decent (SGD), we propose to use the union (upper) bound on the BER of the MRBs as a loss function. By simulations of a multi-span wavelength division multiplexing (WDM) optical fiber channel, we demonstrate that the proposed MLC can

Toshiki Matsumine was with the Department of Photonics Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, and is now with the Institute of Advanced Sciences, Yokohama National University, Yokohama, Japan, email: matsumine-toshiki-mh@ynu.ac.jp

Metodi P. Yankov and Søren Forchhammer are with the Department of Photonics Engineering, Technical University of Denmark, Kongens Lyngby, Denmark, email: {meya, sofo}@fotonik.dtu.dk

This work was supported by the IFD INCOM project (no. 8057-00059B).

achieve comparable or even superior error rate performance compared to the conventional BICM with much lower inner decoding complexity. We also demonstrate that the use of the proposed loss function results in better error rate performance than that of the conventional binary cross entropy loss. We also confirm the effectiveness of our design GS based on simulations using the split step Fourier method (SSFM).

The rest of this paper is organized as follows: Section II introduces the system model of the two-level MLC that we consider throughout this paper. In Section III, we introduce the proposed learning-based approach to GS, and propose a new loss function for learning GS. Section IV verifies by simulations the superiority of the proposed loss function to the conventional loss function based on the binary cross entropy. Furthermore, we demonstrate that the proposed MLC with GS achieves better performance-complexity trade-offs than the conventional BICM with GS. Finally, Section V concludes this work.

II. Two-Level Multi-Level Codes (MLC) with Multi-Stage Decoding (MSD)

In this section, we introduce the two-level MLC system (without GS) employing quadrature amplitude modulation (QAM).

A. The System Model

The overall system model of the proposed two-level MLC with $M(=2^m)$-ary QAM is dipicted in Fig. 1, where only the $l \in \mathbb{Z}$ $(0 < l \leq m)$ LRBs are protected by an inner code. Note that $l = m$ correspond to the conventional BICM system. In this paper, we consider l to be integer, i.e., the number of inner coded bit levels is fixed over the entire transmission block, while it is also possible to mix different numbers of inner coded bit levels [24], [26], [48].

1) Transmitter: As illustrated in Fig. 1, information bits are encoded by the outer FEC whose output is interleaved by a bit interleaver π_1. Then the $m-l$ MRBs are directly fed into the symbol mapper, whereas the l LRBs are input to the inner encoder. After applying bit interleaveing π_2, the codeword bits of the inner FEC are mapped onto $M(=2^m)$-ary QAM together with the MRBs.

Let $\mathbf{B} = (\mathbf{B}_{\text{LRB}}, \mathbf{B}_{\text{MRB}}) = (B_0, \ldots, B_{m-1}) \in \mathbb{F}_2^m$ denote mapper input bits that form a single QAM symbol, where $\mathbf{B}_{\text{LRB}} = (B_0, \ldots, B_l) \in \mathbb{F}_2^l$ and $\mathbf{B}_{\text{MRB}} = (B_{l+1}, \ldots, B_{m-1}) \in \mathbb{F}_2^{m-l}$ denote the l LRBs and $m-l$ MRBs, respectively. Letting $f_{\text{mapper}} : \mathbb{F}_2^m \to \mathbb{C}$ denote a symbol mapping function, a transmit symbol is then expressed as $X = f_{\text{mapper}}(\mathbf{B}) \in \mathcal{X}$, where $\mathcal{X}$ denote a set of constellation points whose cardinality is $|\mathcal{X}| = M$.

Letting R_{outer} and R_{inner} denote the code rates of the outer and inner FEC codes, the overall information rate of the proposed system is given by

$$R = R_{\text{outer}}((m-l) + lR_{\text{inner}}) \text{ [bits/QAM symbol]} \quad (1)$$

2) Nonlinear Channel Model: We consider the NLIN model [47] that takes into account modulation dependent effects and allows accurate analysis of GS constellations, whereas the GN-model [49] assumes the nonlinear interference is a memoryless AWGN term that depends only on the launch power per channel.

For the NLIN model [47], the effective SNR is defined as

$$\text{SNR}_{\text{eff}} = \frac{P_{\text{tx}}}{\sigma^2_{\text{ASE}} + \sigma^2_{\text{NLIN}}} \quad (2)$$

where P_{tx} is the optical launch power, σ^2_{ASE} is the amplified spontaneous emission (ASE) noise from the optical amplifiers, and σ^2_{NLIN} is the NLI variance that includes both intra- and inter-channel distortions. The NLI variance is a function of the fourth and sixth order moments of the constellation, and thus modulation-dependent.

A received symbol may be modeled as

$$Y = X + \underbrace{Z_{\text{ASE}} + Z_{\text{NLIN}}}_{\triangleq Z}, \quad (3)$$

where Z_{ASE} and Z_{NLIN} are independent zero-mean Gaussian random variables with variances σ^2_{ASE} and σ^2_{NLIN}, respectively, so that the sum $Z \triangleq Z_{\text{ASE}} + Z_{\text{NLIN}}$ is a zero-mean Gaussian random variable with variance $\sigma^2 \triangleq \sigma^2_{\text{ASE}} + \sigma^2_{\text{NLIN}}$. With this channel model, the transition probability of the channel is given by

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|y-x|^2}{2\sigma^2}\right). \quad (4)$$

The NLI variance σ^2_{NLI} is modeled as a function of the fourth and sixth moments of a channel input. In general, larger fourth and sixth moments result in larger nonlinearities, and they are maximized by complex continuous Gaussian signaling, while minimized by a constant-modulus modulation such as phase-shift keying (PSK). Constellation shaping typically increases the fourth and sixth moments over uniform QAM as it generates Gaussian-like symbol distribution.

3) Receiver: At the receiver side, multi-stage decoding (MSD) is employed, where we first decode the l LRBs by inner SD decoding, and then the remaining bit levels are estimated by hard-decision conditioned on the estimates of the LRBs.

For the LRBs, the log-likelihood ratio (LLR) of the kth mapper input B_k with $k \in \{0, \ldots, l-1\}$, is calculated as

$$\begin{aligned} L(B_k) &\triangleq \log \frac{P_{B_k|Y}(B_k = 0|y)}{P_{B_k|Y}(B_k = 1|y)} \\ &= \frac{\sum_{x \in \mathcal{X}(B_k=0)} P_X(x) f_{Y|X}(y|x)}{\sum_{x \in \mathcal{X}(B_k=1)} P_X(x) f_{Y|X}(y|x)} \end{aligned} \quad (5)$$

where $\mathcal{X}(B_k = b_k)$ denotes a set of constellation points whose kth bit label is $b_k \in \{0, 1\}$, and $P_X(x) = 1/M$ for $\forall x \in \mathcal{X}$ in this work. After inner SD decoding, the estimated inner codeword is passed to the hard decision block after applying the bit interleaving π_2, which corresponds to the estimates of the LRBs $\widehat{\mathbf{B}}_{\text{LRB}}$. In the hard decision block, the estimate on the MRBs, denoted by $\widehat{\mathbf{B}}_{\text{MRB}}$, is obtained based on $\widehat{\mathbf{B}}_{\text{LRB}}$, as

$$\widehat{\mathbf{B}}_{\text{MRB}} = \underset{\mathbf{B}_{\text{MRB}} \in \mathbb{F}_2^{m-l}}{\arg\min} |Y - f_{\text{mapper}}(\widehat{\mathbf{B}}_{\text{LRB}}, \mathbf{B}_{\text{MRB}})|. \quad (6)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

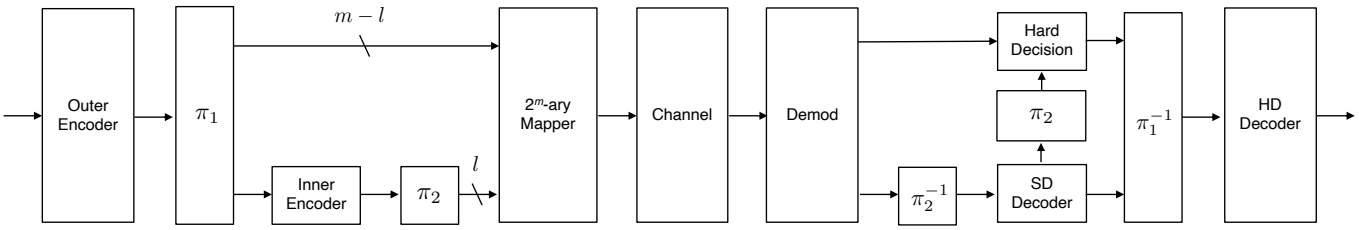


Fig. 1: The system model of concatenated two-level MLC with QAM.

where $|\cdot|$ is the absoute value. The estimated MRBs are then fed into the outer HD decoder together with the estimates of the LRBs after deinterleaving function π_1^{-1}.

In general, two-level MLC performs well in SNR regions where the inner SD decoding starts succeeding, i.e., the inner decoder outputs no bit errors. For high SNR regions where the LRBs are decoded correctly, the BER before outer decoding, which we refer to as pre-FEC BER in this paper, is dominated by the BER of the MRBs. When the number of inner coded bits l is small relative to m, two-level MLC may exhibit an error floor due to the inner uncoded MRBs. However, when concatenated with an outer HD decoder, this would not be problematic as long as the error floor appears below the BER threshold of the outer HD decoder and a sufficiently long random interleaver is available.

B. An Achievable Information Rate (AIR)

An AIR represents the number of bits per symbol that can be transmitted with an arbitrary small error under the assumption of an infinitely long codeword and the chosen system architecture, and thus a relevant performance metric to be maximized. For standard BICM combined with a binary SD FEC, the GMI represents an AIR of the system with a bit-metric decoding architecture. Assuming the Gaussian channel model and the optimal maximum likelihood (ML) detection based on (5), the GMI corresponding to the LRBs of the MLC architecture in Fig. 1 is computed via Monte-Carlo integration as [50]

$$\begin{aligned} &R_{\text{GMI}}(\mathbf{B}_{\text{LRB}}, L(\mathbf{B}_{\text{LRB}})) \\ &\quad \triangleq l - \sum_{k=0}^{l-1} \mathbb{E}_{B_k, L(B_k)} \left[\log_2 (1 + e^{(-1)^{B_k} L(B_k)}) \right], \end{aligned} \tag{7}$$

where $\mathbb{E}_{B_k, L(B_k)}$ is the expectation with respect to B_k and $L(B_k)$, and the LRBs follow uniform distribution, i.e., $P(\mathbf{B}_{\text{LRB}}) = (1/2)^l$.

For the proposed system in Fig. 1, we may consider an AIR in the following two scenarios. First, we consider the ideal case where the inner code rate is matched to the underlying channel condition, i.e., $R_{\text{inner}} = R_{\text{GMI}}/l$, and the corresponding decoder outputs no bit error. In this case, we may separately protect the MRBs and the LRBs by outer and inner codes, respectively. Denoting the BER at the outer decoder input by ϵ_{MRB}, an AIR of the entire system, denoted by $R^{\star}_{\text{TL-MLC}}$, is given as

$$R^{\star}_{\text{TL-MLC}} = (m-l)(1 - \mathbb{H}_{\text{b}}(\epsilon_{\text{MRB}})) + R_{\text{GMI}}, \tag{8}$$

where $\mathbb{H}_{\text{b}}(\cdot)$ is the binary entropy function. The first and the second terms of (8) correspond to AIRs of the MRBs under binary hard decision decoding and the LRBs under soft decision decoding, respectively.

We next consider the case of pragmatic inner coding that results in a non-zero BER after inner decoding. In this case, the BER at the input of the outer decoder, denoted by ϵ_{pre}, includes bit errors from both the MRBs and the LRBs. An AIR of this system is given by

$$R_{\text{TL-MLC}} = (m-l) + lR_{\text{inner}} - m\mathbb{H}_{\text{b}}(\epsilon_{\text{pre}}), \tag{9}$$

where the first two terms correspond to the maximum achievable rate without channel uncertainty, i.e., without noise, and the third term represents the uncertainty that the receiver has.

Note that since the BER after inner decoding as well as ϵ_{pre} depend on the specific inner code we employ, maximization of (9) also depends on performance of the inner code. In this work, since we don't assume any specific inner coding scheme, we consider the first case and attempt to maximize (8), i.e., we fix R_{inner} to R_{GMI} and assume it is not a subject to optimization in order to reduce the degrees of freedom in the optimization procedure. We note that the performance can potentially be improved by including R_{inner} in the optimization procedure, however, we expect the gains to be small.

C. Bit-Labeling

Appropriate bit-labeling design is necessary to achieve good error rate performance for the proposed two-level MLC. More specifically, we wish to maximize the GMI for the LRBs, while minimizing the BER of the MRBs under MSD. In what follows, we describe how to systematically construct a bit-labeling scheme for QAM to meet these requirements. Here we consider one-dimensional PAM for simplicity of explanation, however, extension to QAM is straightforward by allocating two PAMs to real and imaginary components of QAM.

Consider a set of $\sqrt{M}(= 2^{m/2})$-ary PAM constellation points $\mathcal{P} = \{\pm 1, \pm 3, \ldots, \pm(\sqrt{M}-1)\}$, which is equivalent to M-ary QAM, where the $l/2$ LRBs are encoded by inner FEC and the $(m-l)/2$ MRBs remain uncoded. The proposed labeling is constructed from the two PAMs corresponding to the MRBs and the LRBs, which are expressed as $\mathcal{P}_{\text{MRB}} = \left\{ -(\sqrt{M}-1) + 2iN_{\text{LRB}} \mid i \in \{0, 1, \ldots, N_{\text{MRB}} - 1\} \right\}$, and $\mathcal{P}_{\text{LRB}} = \{2i \mid i \in \{0, 1, \ldots, N_{\text{LRB}} - 1\}\}$, respectively, where $N_{\text{MRB}} = 2^{(m-l)/2}$ and $N_{\text{LRB}} = 2^{l/2}$. We use Gray labeling for both $\mathcal{P}_{\text{MRB}}$ and $\mathcal{P}_{\text{LRB}}$ to minimize the BER of the MRBs and maximize the GMI of the LRBs [51], respectively. Then,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

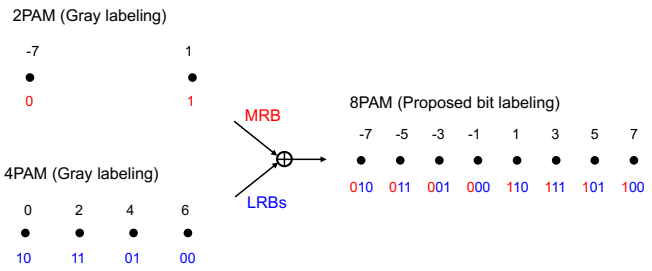


Fig. 2: Example of the proposed bit-labeling for two-level MLC with 8PAM and two inner coded bit levels ($m = 6$ and $l = 4$).

a PAM symbol $x \in \mathcal{P}$ is constructed by the superposition of the two PAMs, i.e., $x = i + j$ where $i \in \mathcal{P}_{\text{MRB}}$ and $j \in \mathcal{P}_{\text{LRB}}$, so that the minimum Euclidean distance is increased from 2 to $2N_{\text{LRB}}$ after decoding the LRBs.

In Fig. 2, we illustrate the procedure of constructing the proposed bit-labeling for 8-ary PAM ($m = 6$ and $l = 4$) from 2PAM $\mathcal{P}_{\text{MRB}} = \{-7, 1\}$ and 4PAM $\mathcal{P}_{\text{LRB}} = \{0, 2, 4, 6\}$ corresponding to the MRB and the LRBs, respectively. Note that while the original minimum Euclidean distance is 2, it is increased to 8 after decoding the LRBs.

D. Complexity Issue

The major advantage of the proposed MLC scheme over the conventional BICM is its significantly reduced complexity associated with inner SD decoding. This reduction of SD decoding complexity may have a significant impact on the overall system complexity as SD decoding is in general more power consuming than HD decoding [52]–[54].

Complexity of two-level MLC that encodes only the LRB has been studied in [4], [5]. In this work, we follow the complexity measures for LDPC-coded MLC and BICM systems introduced in [5], which are expressed as

$$\eta_{\text{MLC}} = \frac{N_{\text{edge}} N_{\text{iter}}}{N((m-l)/2 + R_{\text{inner}})}, \quad \eta_{\text{BICM}} = \frac{N_{\text{edge}} N_{\text{iter}}}{N R_{\text{inner}}}, \tag{10}$$

where N, N_{iter}, and N_{edge} are the code length, the average number of decoding iterations, and the number of edges in the graph of the code, respectively. Note that (10) accounts for the fact that the proposed MLC encodes only $l/2$ bits per $2^{m/2}$-ary PAM by an inner code. As shown in (10), the overall decoding complexity of LDPC codes is proportional to the numbers of decoding iterations and edges in the graph [55], [56].

III. Neural Network-Based Geometric Shaping for Two-Level MLC

In this section, we introduce the proposed approach to training a symbol mapper for the proposed two-level MLC. We first describe the structure of the proposed symbol mapper, and then elaborate on the proposed loss function suited for the proposed MLC.

A. Neural Network-Based Symbol Mapper

We illustrate a schematic diagram for the proposed learning approach in Fig. 3. In this paper, we consider a neural network (NN) with no hidden layer as an encoder [40] where the weights directly correspond to the coordinates of constellation points. Letting $\mathbf{B} \in \mathbb{F}_2^m$ denote an input bit sequence to the symbol mapper, an n-dimensional transmit symbol is then expressed as

$$X = \mathbf{W}\phi_{\text{onehot}}(\mathbf{B}) \tag{11}$$

where $\phi_{\text{onehot}} : \mathbb{F}_2^m \to \mathbb{F}_2^M$ is a one-hot encoding function, and $\mathbf{W} \in \mathbb{R}^{n \times M}$ is the weights of the neural network. In this work, we exclusively focus on the case with $n = 2$, where the first and second elements correspond to real and imaginary parts of the complex plane.

At the receiver side, we employ ML detection based on (5) assuming the Gaussian channel model, unlike a conventional autoencoder that employs NN at the receiver. It is worthwhile mentioning that, while the symbol mapper has a single layer, the demapper is a function of the symbol mapper, and thus the proposed system may be seen as a neural network with multiple layers. In this sense, the proposed system is related to the concept of autoencoder [34], [35], [37], [40], [41], [43]–[45] that jointly trains transmitter and receiver in an end-to-end manner.

As mentioned above, the symbol mapper has $2 \times W$ weights (ignoring a bias term) to be trained. These weights can be initialized with either random Gaussian variables or specific values. In fact, it has been already known that domain-specific knowledge may be helpful [40] in learning communication signals. More specifically, it was shown that initializing the neural network weights with Gray-labeled QAM, which is known to be good for BICM systems, can accelerate the learning process compared to a random initialization. Therefore, we may use QAM with the bit-labeling described in Section II-C as an initialization for the proposed two-level MLC. Note that, the problem is highly nonconvex, and thus it will be cumbersome to find a global optimum as mentioned in [40].

B. The Proposed Learning Objective

Our primary objective of training is to maximize the AIR (8). We see that (8) will be maximized by maximizing the GMI for the LRBs which will minimize the BER after SD decoding and also the BER of the MRBs after MSD. The BER of the MRBs will be further reduced by proper symbol mapping and constellation designs.

1) Loss Function for the LRBs: In the proposed MLC, the l least reliable bit levels encoded by an inner code, are similar to BICM. Therefore, the appropriate performance metric for the LRBs would be GMI. It has been addressed in [57]–[59] that minimization of the binary cross entropy loss is equivalent to maximization of the GMI (7). Therefore, we use the binary cross entropy loss for the LRBs. Let $\mathcal{J}_{\text{BCE}}(x, y)$ denote the binary cross entropy loss function between x and y, and $\phi_{\text{sigmoid}}(x) = 1/(1 + e^{-x})$ denote the sigmoid function. Also,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

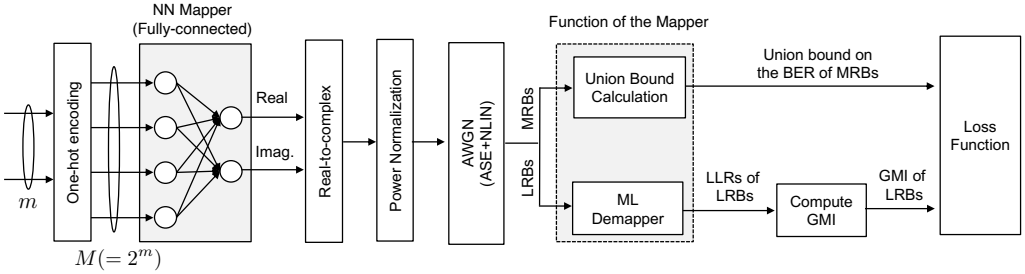


Fig. 3: The system model of concatenated two-level MLC.

letting $\Phi(B_k) \triangleq \phi_{\text{sigmoid}}(L(B_k))$, the binary cross entropy between $\mathbf{B}_{\text{LRB}}$ and $\Phi(\mathbf{B}_{\text{LRB}})$ is written as

$$\begin{aligned}&\mathcal{J}_{\text{BCE}}(\mathbf{B}_{\text{LRB}}; \Phi(\mathbf{B}_{\text{LRB}}))\\ &\quad= \frac{1}{l}\sum_{k=0}^{l-1} \mathbb{E}_{B_k, L(B_k)}\left[\log_2(1+e^{(-1)^{B_k}L(B_k)})\right]\\ &\quad= 1 - R_{\text{GMI}}(\mathbf{B}_{\text{LRB}}, L(\mathbf{B}_{\text{LRB}}))/l. \end{aligned} \tag{12}$$

2) Loss Function for the MRBs: Again, while we wish to minimize the BER of the MRBs, the problem of the BER as a loss function is its nondifferentiability. To circumvent this issue, we apply the union upper bound on the BER as a loss function to be minimized, which is differentiable.

Let $\mathcal{X}(\mathbf{B}_{\text{LRB}})$ denote a set of constellation points corresponding to the LRBs $\mathbf{B}_{\text{LRB}}$ whose cardinality is given by $|\mathcal{X}(\mathbf{B}_{\text{LRB}})| = 2^{m-l}$. We assume that the LRBs $\mathbf{B}_{\text{LRB}} \in \mathbb{F}_2^l$ are uniformly distributed so we have $P(\mathbf{B}_{\text{LRB}}) = (1/2)^l$. The BER of the MRBs is then upper bounded by the union bound $\mathcal{J}_{\text{UB}}$, which is expressed as,

$$\mathcal{J}_{\text{UB}} = \sum_{\mathbf{B}_{\text{LRB}} \in \mathbb{F}_2^l} P(\mathbf{B}_{\text{LRB}}) \sum_{\substack{x, \widehat{x} \in \mathcal{X}(\mathbf{B}_{\text{LRB}}) \\ x \neq \widehat{x}}} \beta(x, \widehat{x})\, P(x \to \widehat{x}) \tag{13}$$

where $x, \widehat{x}$ denote the transmitted and estimated symbols at the receiver side, respectively, and $P(x \to \widehat{x})$ is the pairwise error probability that the transmitted symbol x is detected as $\widehat{x}$ at the receiver side. For an AWGN channel with noise variance σ^2, it is expressed as

$$P(x \to \widehat{x}) = Q\left(\sqrt{\frac{|x-\widehat{x}|^2}{4\sigma^2}}\right), \tag{14}$$

where $Q(x) = \frac{1}{2}\text{erfc}(\frac{x}{\sqrt{2}}) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-\frac{u^2}{2})du$ is the Q-function. Also, we define the coefficient that represents the ratio of incorrect bits in the estimated MRBs at the receiver side as

$$\beta(x, \widehat{x}) \triangleq \frac{d_{\text{H}}(\mathcal{B}_{\text{MRB}}(x), \mathcal{B}_{\text{MRB}}(\widehat{x}))}{m-l}, \tag{15}$$

where $d_{\text{H}}(\cdot,\cdot)$ denotes the Hamming distance between two bit sequences, and $\mathcal{B}_{\text{MRB}}(x) \in \mathbb{F}_2^{m-l}$ denotes the inverse symbol mapping function that maps a modulation symbol into the MRBs

The union bound will become tight for high SNR regions where the LRBs are decoded correctly by the inner decoder, and the pre-FEC BER is dominated by the BER of the MRBs. Note that the Gray labeling of the MRBs is important to minimize the term $\beta(x, \widehat{x})$. Also, for a high SNR region, the minimization of the union bound (13) is similar to maximization of the minimum Euclidean distance since the effect of the minimum Euclidean distance on the union bound becomes dominant. The union bound described above is differentiable since the complementary error function (erfc) function is differentiable as $\frac{d}{dx}\text{erfc}(x) = \frac{-2}{\sqrt{\pi}}e^{-x^2}$. We also note that the calculation of $\mathcal{L}_{\text{UB}}$ does not require simulations for the receiver as it depends only on the symbol mapper and the channel.

3) Overall Loss Function: We wish to simultaneously minimize the aforementioned two different loss functions for the MRBs and the LRBs, respectively, which is cast as multi-objective optimization. In this work, we consider the following loss function that combines the two loss functions with scaling;

$$\mathcal{J}_{\text{Loss}} = \alpha \mathcal{J}_{\text{BCE}}(\mathbf{B}_{\text{LRB}}; \Phi(\mathbf{B}_{\text{LRB}})) + \mathcal{J}_{\text{UB}} \tag{16}$$

where $\alpha \in \mathbb{R}_+$ is a (non-negative) parameter to control the balance of the two loss functions. It is expected that, if we put a higher priority on the LRBs (a higher value of α), performance will be improved in a low SNR region, whereas performance will be degraded in a high SNR region. In contrast, performance in a high SNR region will be improved by decreasing α. As such, the proposed system provides a wide range of trade-offs between waterfall and error floor performances by adjusting the parameter α depending on a required pre-FEC BER.

In the proposed optimization, the parameter α has to be empirically chosen depending on an acceptable error floor. Therefore, we may perform optimization for various values of α and thus the proposed optimization may require more computational effort and optimization time than the conventional cross-entropy function. However, we note that the optimization is performed *offline*, and this does not affect real-time communications.

IV. Numerical Results

In this section, we evaluate the pre-FEC BER performance of the proposed GS in comparison with the conventional BICM with GS. For BICM with GS, the symbol mapper is trained based on the standard binary cross entropy minimization.

For the outer code, we consider two different codes: the (30832, 30592) BCH codes with a code rate of 0.992 whose BER threshold is 5.0×10^{-5} [60], [61], and the staircase code

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

TABLE I: The channel parameters used in simulations.

Symbol rate	32 GHz
Oversampling	32
Channel Spacing	50 GHz
The number of channels	5
The number of polarizations	2
Pulse shaping	Root raise cosine (roll-off: 0.05)
Span length	100 km
Nonlinear coefficient	1.3 $(\text{W km})^{-1}$
Dispersion parameter	16.48 ps/(nm km)
Attenuation	0.2 dB/km
EDFA noise figure	5 dB
SSFM step size	0.1 km

with a code rate of $239/255 = 0.937$ whose BER threshold is 5.0×10^{-3} [62], [63]. As an inner code, we employ LDPC codes of length 64800 bits from the DVB-S2 standard [64]. We use belief propagation (BP) decoding with the maximum iteration count of 50 for LDPC codes. However, we would like to emphasize that any combination of outer and inner codes is applicable to the proposed MLC. For the interleaver π_2 between the inner code and the symbol mapper, we assume a random permutation.

For a fair comparison, the performances of the proposed MLC and BICM are compared with the same data rate. To do so, we use different inner code rates for MLC and BICM, while the code lengths are the same. Note that even with the same inner code length, the proposed MLC and the conventional BICM result in different symbol lengths. More specifically, the resulting transmit symbol lengths for the proposed MLC with l SD coded bit levels and the conventional BICM are $64800/l$ and $64800/m$ complex symbols, respectively.

A. Channel and Training Parameters

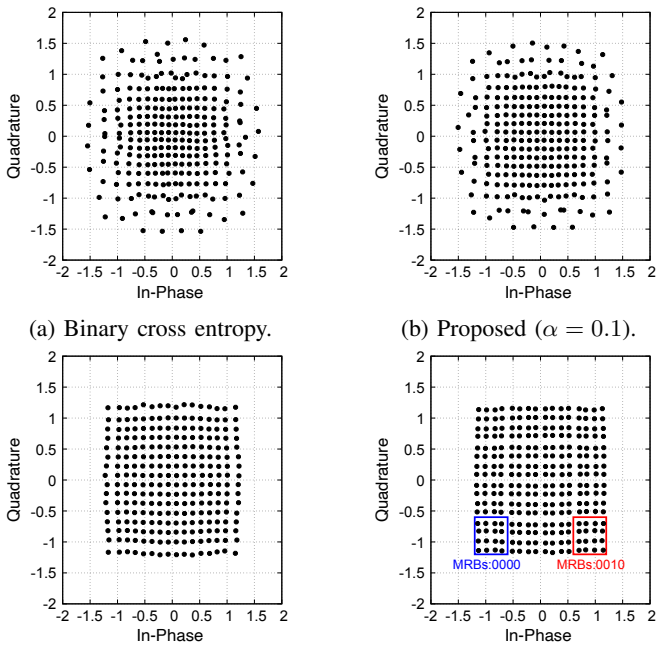
We simulate a multi-span optical fiber system based on the NLIN model [47] and an actual SSFM with channel parameters given in Table I.

We use Tensorflow [65] for implementing the proposed communication system. The training of the symbol mapper in the proposed system involves several hyperparameters. We use Adam optimization with a learning rate 0.001 for optimization. The training datasets consists of approximately 10^6 bits, and training is performed with a batch size of 1024 bits for 100 epochs, which is sufficient for convergence. Instead of training at a specific optical launch power, we let the launch power be a training parameter, and optimize it jointly with the symbol mapper as in [66].

B. Effect of the Training Parameter α

In what follows, we assess the effect of the training parameter α on performance of the proposed system.

1) Learned Constellation: We first plot in Fig. 4, the learned 256-ary constellations for $l = 4$ at a transmission distance of 1000 km with different training losses. It is observed from this figure that a higher value of α results in Gaussian-like distribution, while for a lower value of α, constellation retains the rectangular shape of the standard square QAM.



(a) Binary cross entropy. (b) Proposed ($\alpha = 0.1$).

(c) Proposed ($\alpha = 0.01$). (d) Proposed ($\alpha = 0.005$).

Fig. 4: Learned 256-ary constellation ($m = 8$) for $l = 4$ at a transmission distance of 1000km with the conventional and the proposed loss functions.

This is because the square QAM is better in terms of the union bound due to its larger minimum Euclidean distance and Gray labeling. Specifically, for $\alpha = 0.005$ (Fig. 4d), it appears that the constellation is divided into 2^4 subgroups corresponding to the possible patterns of 4-bit LRBs (constellations in each subgroup have the same MRB label), which would maximize the minimum Euclidean distance after knowing the LRBs. Also, the optimized constellation with $\alpha = 0.1$ (Fig. 4b) visually looks similar to the constellation optimized using the conventional binary cross entropy (Fig. 4a). However, we show in the sequel that the our proposed constellation (Fig. 4b) can actually achieve better BER performance than the conventional one (Fig. 4a).

2) Pre-FEC BER with LDPC Codes: Figure 5 shows the pre-FEC BER of the proposed MLC with 256-ary constellation learned using the conventional binary cross entropy loss and the proposed loss function (16) over the NLIN channel model. In this figure, we use the rate-$1/4$, -$1/2$ codes in DVB-S2 for $l = 4$ and $l = 6$, respectively, such that a data rate is fixed to 5 bits per complex symbol ignoring the outer FEC code. Performances over the AWGN channel for the same simulation setting are presented in Fig. 6. The BER thresholds of the selected outer BCH and staircase codes are indicated by horizontal lines.

We expect from (16) that a lower value of α results in a lower error floor since higher priority is put on the union bound for the MRBs. This is confirmed from Fig. 5 and Fig. 6, where we observe that the proposed loss function offers a wide range of trade-offs between performances in the waterfall and the error floor regions by changing a parameter α. Specifically, we observe from Fig. 5 that the proposed loss

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

function with $\alpha = 0.1$ achieves the BER threshold of the outer staircase code whereas the conventional cross entropy loss does not achieve the threshold even at the optimal launch power (around 0–0.5 dBm). This result demonstrates that the proposed loss function can offer a reach increase with respect to the conventional cross entropy function.

We also observe from Fig. 6 that, the error floor of the proposed MLC can be significantly reduced by increasing the number of inner coded bit levels l. Specifically, for $l = 6$, error floor is not visible above a BER of 10^{-5} (Our union bound analysis suggests that the error floor in this case will appear well below 10^{-6}, which is much lower than typical outer BER thresholds). On the other hand, it is observed that increasing l degrades performance at waterfall regions. A larger value of l results in higher decoding complexity from the discussion in Section II-D. Therefore, when powerful outer HD codes are available, e.g., staircase codes [63], the proposed two-level MLC with a fewer inner coded bit levels, i.e., $l = 4$, may be more effective, while $l = 6$ is required for relatively weak outer HD codes, i.e., a single BCH code [60], [61].

In practice, GS requires a finer digital-to-analog converter (DAC) and analog-to-digital converter (ADC) precision than PS due to its irregular constellation points and this drawback of GS can limit its shaping gain. Therefore, we further investigate the quantization effect of finite precision ADC and DAC on the performance of the proposed scheme. We assume an ideal frequency-flat quantization noise, which is modeled as a random variable following uniform distribution $\mathcal{U}(-D/2^{Q-1}, D/2^{Q-1})$, where Q is the number of quantization bits and D is the dynamic range of the ADC and DAC. We assume $D = 1.2 \max_{X \in \mathcal{X}} \mathrm{Re}(X)$ in this work where $\mathrm{Re}(X)$ is the real part of X. The quantization noise is injected at both transmitter and receiver.

In Fig. 7, we evaluate based on an SSFM the performances of the conventional BICM with GS, and the proposed MLC with GS using the proposed and the conventional loss functions. In these simulations, we also consider finite quantization precision of Q bits as described above. Simulation parameters are the same as Fig. 5. From this figure, it is observed that the proposed MLC with GS outperforms the conventional BICM in terms of the pre-FEC BER. This result verifies that the proposed MLC with GS is effective not only for the NLIN model but also for an actual SSFM simulation. Also, we observe that $Q = 5$ bits may not be sufficient to see the gain of the proposed loss function over the conventional one. However, with $Q = 6$ bits, the proposed loss function clearly outperforms the conventional one at a launch power of around 0 dBm, and achieves the outer BER threshold. This demonstrates that the gains achieved by the proposed loss function will be meaningful and reproducible with practical quantization precision.

C. Rate vs. Transmission Distance

In what follows, we evaluate the achievable rate with the practical concatenated FEC scheme of the outer staircase and the inner DVB-S2 LDPC code. The achievable rate is given as the maximum information rate where the pre-FEC BER

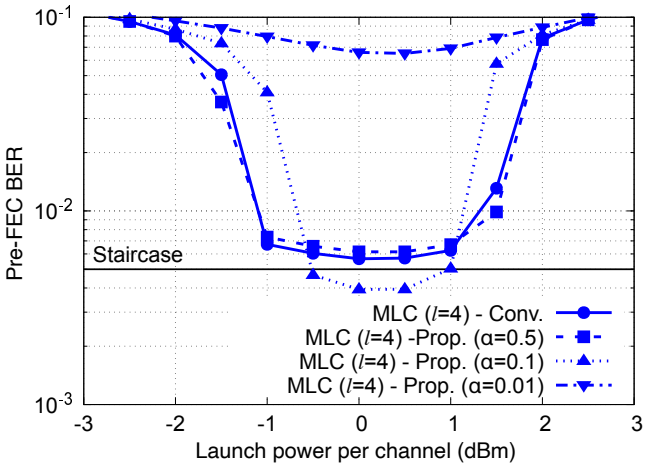


Fig. 5: Pre-FEC BER performances of MLC with the conventional and the proposed loss functions for the NLIN channel model. 256-ary modulation is employed and a data rate is set to 5 bits per symbol without outer codes. A transmission distance is 1500km.

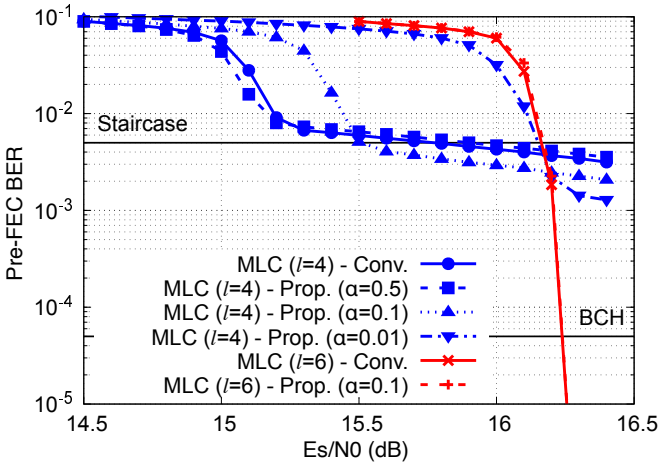


Fig. 6: Pre-FEC BER performances of MLC with the conventional and the proposed loss functions for the AWGN channel. 256-ary modulation is employed and a data rate is set to 5 bits per symbol without outer codes.

reaches an outer threshold, and thus the target system BER of 10^{-15} is achieved assuming an ideal bit interleaving between the outer and inner codes. We always use constellations optimized for a transmission distance of 1000 km with a training parameter $\alpha = 0.1$, and various information rates are realized for a fixed l by adjusting the code rate of the inner LDPC code. Note that while our GS optimization does not impose any constraint on outer and inner code rates, there are target operating code rates in practice. Therefore, there may be room for performance improvement by including them in the optimization procedure. In Table II and Table III, the optimized launch powers and AIRs (8) with the optimized GS are summarized, respectively.

In the following simulations, the transmission distance is varied between 500 and 3000 km in steps of 100 km. Although

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

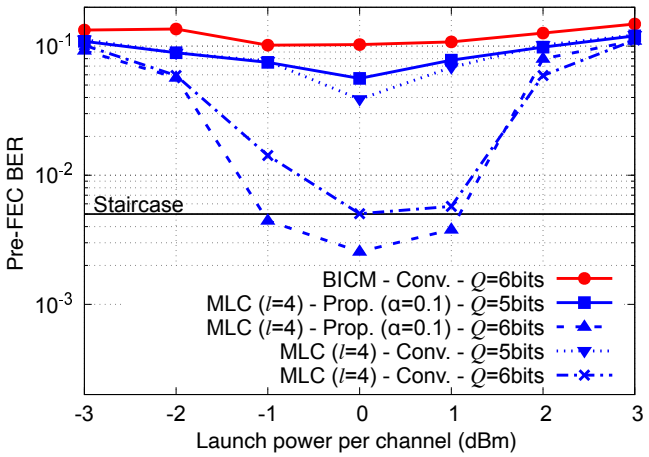


Fig. 7: Pre-FEC BER performances with an SSFM and Q quantization bits at a transmission distance 1300km. 256-ary modulation is employed and a data rate is set to 5 bits per symbol without outer codes.

TABLE II: The optimized launch powers per channel in dBm.

	BICM	MLC ($l=2$)	MLC ($l=4$)	MLC ($l=6$)
64QAM	0.41	0.43	0.42	n/a
256QAM	0.26	n/a	0.36	0.28

the launch power is optimized for the case with GS in the training phase, the optimal launch power is unknown for QAM without GS. Therefore, in order to make fair comparisons between the system with and without GS, we sweep the launch power with a granularity of 0.5 dB and the performance is measured at the optimal power for both cases.

We summarize in Table IV the parameters that determine the decoding complexity of the system, including the number of edges in the code graph N_{edge}, the average number of decoding iterations N_{iter}, and the complexity score η in (10) for some operating rates.

In Fig. 8a, we plot the performance with the 64-ary constellation. It is observed that the proposed MLC with $l = 2$ achieves the best performance at a moderate transmission distance of around 1500–2000 km with much lower inner SD decoding complexity than the conventional BICM. Specifically, at a fixed rate of 9 bits per 4D symbol, the proposed MLC with $l = 2$ offers a reach increase of 100 km over the BICM with GS. At this point, the complexity scores in (10) for the proposed MLC and the conventional BICM are summarized in Table IV. From this Table, we see that the proposed MLC with $l = 2$ achieves about 85 % complexity reduction over the conventional BICM.

For a longer transmission distance, the proposed MLC with $l = 4$ achieves the best performance. Specifically, at a rate of 7.5 bits per 4D symbol, the proposed MLC with $l = 4$ provides a reach increase of 100 km over the conventional BICM. Also from Table IV, we observe that the proposed MLC with $l = 4$ in this case achieves about 64 % complexity reduction over the conventional BICM, in addition to its superior performance in terms of transmission distance.

Figure 8b shows performance with 256-ary constellation,

TABLE III: AIRs (8) with the GS optimized at a transmission distance of 1000 km. The values in parentheses are AIRs of LRBs and MRBs, respectively.

	MLC ($l=2$)	MLC ($l=4$)	MLC ($l=6$)
64QAM	5.39 (1.42, 3.97)	5.41 (3.41, 2.00)	n/a
256QAM	n/a	5.63 (1.67, 3.95)	5.65 (3.65, 2.00)

i.e., 8 bits per complex symbol. With the concatenated staircase and LDPC codes, the proposed MLC with $l = 4$ achieves the best performance at a transmission distance less than 1500 km. Specifically, the proposed MLC with $l = 4$ achieves a reach increase of 200 km over the conventional BICM at an information rate of 10 bits per 4D symbol. Note that in this case, the proposed MLC achieves 70 % complexity reduction in terms of (10) with respect to the conventional BICM.

For a longer transmission distance, e.g., more than 1500 km, the proposed MLC with $l = 6$ achieves the best performance. Specifically, by GS, the transmission distance of the proposed MLC with $l = 4$ is increased from 2200 km to 2600 km. This transmission distance is 300 km longer than that of the conventional BICM with GS, despite 50 % lower SD decoding complexity of the proposed MLC from Table IV. More specifically, the complexity score (10) of the proposed MLC with $l = 6$ is approximately half of that of the conventional BICM. These results demonstrate the superior performance of the proposed MLC over the conventional BICM in terms of performance-complexity trade-offs.

V. Conclusion

In this paper, we have considered a learning approach to constellation optimization for a concatenated two-level MLC scheme. Assuming the NLIN model, we have proposed a new learning objective function for the two-level MLC that combines the generalized mutual information (GMI) for the least reliable bits (LRBs) and the union bound on BER for the most reliable bits (MRBs) in order to minimize the pre-FEC BER, and thus maximize the achievable rate of the proposed system. It was demonstrated by simulations that the proposed loss function offers a performance gain over the conventional cross entropy loss function. Furthermore, we have demonstrated that the proposed MLC can simultaneously offer a performance gain and complexity reduction over the conventional BICM. Specifically, for 256-ary modulation, the proposed MLC offers a reach increase of 300 km even with 50 % lower inner decoding complexity, compared to the conventional BICM. The performance gain by the use of the proposed loss function comes at no cost with respect to the cross entropy loss, i.e., it affects neither transmitter nor receiver complexities in practice.

References

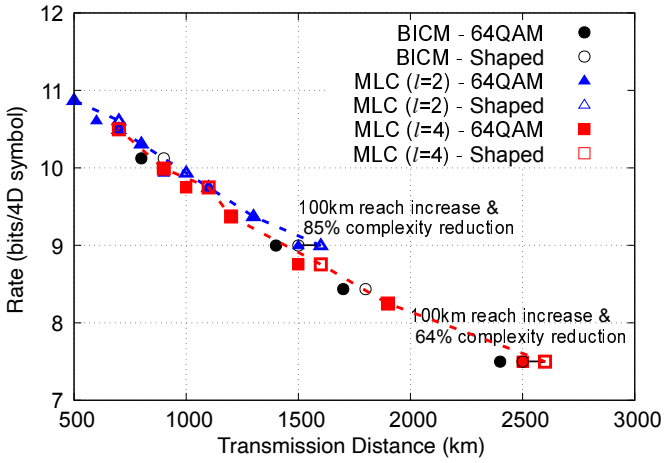
[1] Optical Internetworking Forum, "Implementation Agreement 400ZR," 2018.

[2] A. Bisplinghoff, S. Langenbach, and T. Kupfer, "Low-power, phase-slip tolerant, multilevel coding for M-QAM," *J. Lightw. Technol.*, vol. 35, no. 4, pp. 1006–1014, 2016.

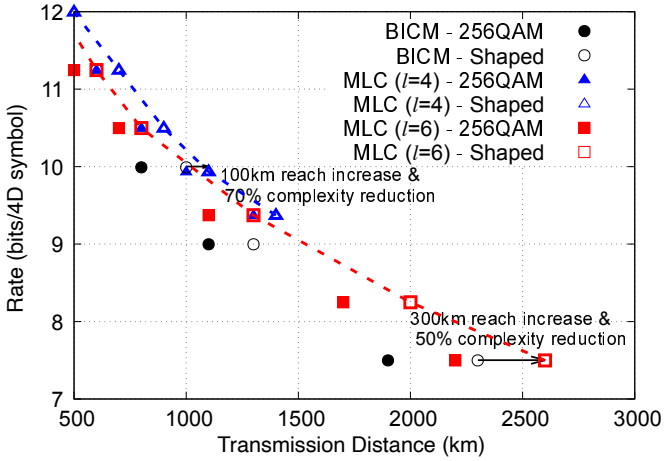
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JLT.2022.3179529, Journal of Lightwave Technology

TABLE IV: The complexity parameters for the proposed MLC and the conventional BICM.

Coding	Modulation	Overall information rate (bits/4D symbol)	R_{inner}	N_{edge}	N_{iter}	η
Proposed MLC	$M = 64$	9	$2/5$ ($l = 2$)	233279	16.4	24.6
		7.5	$1/2$ ($l = 4$)	226799	20.4	47.6
	$M = 256$	10	$1/3$ ($l = 4$)	215999	23.6	33.8
		7.5	$1/3$ ($l = 6$)	215999	30.0	75.0
Conventional BICM	$M = 64$	9	$4/5$	233279	33.7	151.6
		7.5	$2/3$	215999	26.5	131.8
	$M = 256$	10	$2/3$	215999	22.5	112.5
		7.5	$1/2$	226799	21.4	149.8



(a) 64-ary constellation (6 bits per complex symbol).



(b) 256-ary constellation (8 bits per complex symbol).

Fig. 8: The achievable rates with the concatenated staircase-LDPC code for the proposed MLC and the conventional BICM over the NLIN channel.

[3] Y. Koganei, T. Oyama, K. Sugitani, H. Nakashima, and T. Hoshida, "Multilevel coding with spatially coupled repeat-accumulate codes for high-order QAM optical transmission," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 486–492, 2019.

[4] M. Barakatain and F. R. Kschischang, "Low-complexity concatenated LDPC-staircase codes," *J. Lightw. Technol.*, vol. 36, no. 12, pp. 2443–2449, 2018.

[5] M. Barakatain, D. Lentner, G. Böecherer, and F. R. Kschischang, "Performance-complexity tradeoffs of concatenated FEC for higher-order modulation," *J. Lightw. Technol.*, vol. 38, no. 11, pp. 2944–2953, 2020.

[6] M. Barakatain and F. R. Kschischang, "Low-complexity rate-and channel-configurable concatenated codes," *J. Lightw. Technol.*, vol. 39, no. 7, pp. 1976–1983, 2020.

[7] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.

[8] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[9] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, 1992.

[10] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.

[11] F. Frey, S. Stern, J. K. Fischer, and R. Fischer, "Two-stage coded modulation for hurwitz constellations in fiber-optical communications," *J. Lightw. Technol.*, 2020.

[12] D. Rohweder, S. Stern, R. F. Fischer, S. Shavgulidze, and J. Freudenberger, "Four-dimensional Hurwitz signal constellations, set partitioning, detection, and multilevel coding," *IEEE Trans. Commun.*, vol. 69, no. 8, pp. 5079–5090, 2021.

[13] M. P. Yankov, D. Zibar, K. J. Larsen, L. P. Christensen, and S. Forchhammer, "Constellation shaping for fiber-optic channels with QAM and high spectral efficiency," *IEEE Photon. Technol. Lett.*, vol. 26, no. 23, pp. 2407–2410, 2014.

[14] T. Fehenberger, A. Alvarado, G. Böcherer, and N. Hanik, "On probabilistic shaping of quadrature amplitude modulation for the nonlinear fiber channel," *J. Lightw. Technol.*, vol. 34, no. 21, pp. 5063–5073, 2016.

[15] F. Buchali, F. Steiner, G. Böcherer, L. Schmalen, P. Schulte, and W. Idler, "Rate adaptation and reach increase by probabilistically shaped 64-QAM: An experimental demonstration," *J. Lightw. Technol.*, vol. 34, no. 7, pp. 1599–1609, 2016.

[16] M. P. Yankov, F. Da Ros, E. P. da Silva, S. Forchhammer, K. J. Larsen, L. K. Oxenløwe, M. Galili, and D. Zibar, "Constellation shaping for WDM systems using 256QAM/1024QAM with probabilistic optimization," *J. Lightw. Technol.*, vol. 34, no. 22, pp. 5146–5156, 2016.

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

[18] G. Böcherer, P. Schulte, and F. Steiner, "Probabilistic shaping and forward error correction for fiber-optic communication systems," *J. Lightw. Technol.*, vol. 37, no. 2, pp. 230–244, 2019.

[19] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.*, vol. 37, no. 6, pp. 1590–1607, 2019.

[20] T. Yoshida, M. Karlsson, and E. Agrell, "Hierarchical distribution matching for probabilistically shaped coded modulation," *Journal of Lightwave Technology*, vol. 37, no. 6, pp. 1579–1589, 2019.

[21] A. Amari, S. Goossens, Y. C. Gültekin, O. Vassilieva, I. Kim, T. Ikeuchi, C. M. Okonkwo, F. M. Willems, and A. Alvarado, "Introducing enumerative sphere shaping for optical communication systems with short blocklengths," *J. Lightw. Technol.*, vol. 37, no. 23, pp. 5926–5936, 2019.

[22] B. Chen, C. Okonkwo, H. Hafermann, and A. Alvarado, "Polarization-ring-switching for nonlinearity-tolerant geometrically shaped four-dimensional formats maximizing generalized mutual information," *J. Lightw. Technol.*, vol. 37, no. 14, pp. 3579–3591, 2019.

[23] B. Chen, A. Alvarado, S. van der Heide, M. van den Hout, H. Hafermann, and C. Okonkwo, "Analysis and experimental demonstration of orthant-symmetric four-dimensional 7 bit/4D-sym modulation for optical fiber communication," *J. Lightw. Technol.*, vol. 39, no. 9, pp. 2737–2753, 2021.

[24] K. Sugitani, Y. Koganei, T. Oyama, and H. Nakashima, "Partial multi-level coding with probabilistic shaping for low-power optical transmission," in *Proc. 24th OptoElectronics and Communications Conference*

(OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC), pp. 1–3, 2019.

[25] T. Yoshida, M. Karlsson, and E. Agrell, "Multilevel coding with flexible probabilistic shaping for rate-adaptive and low-power optical communications," in *Proc. Optical Fiber Communications Conference (OFC)*, pp. 1–3, 2020.

[26] K. Sugitani, Y. Koganei, H. Irie, and H. Nakashima, "Performance evaluation of WDM channel transmission for probabilistic shaping with partial multilevel coding," *J. Lightw. Technol.*, vol. 39, no. 9, pp. 2873–2879, 2021.

[27] T. Matsumine, M. P. Yankov, T. Mehmood, and S. Forchhammer, "Rate-adaptive concatenated multi-level coding with novel probabilistic amplitude shaping," *IEEE Trans. Commun.*, vol. 70, no. 5, pp. 2977–2991, 2022.

[28] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth efficient and rate-matched low-density parity-check coded modulation," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4651–4665, Dec. 2015.

[29] F. Steiner and G. Böcherer, "Comparison of geometric and probabilistic shaping with application to ATSC 3.0," in *11th International ITG Conference on Systems, Communications and Coding (SCC)*, pp. 1–6, 2017.

[30] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Physical layer deep learning of encodings for the MIMO fading channel," in *Proc. 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 76–80, 2017.

[31] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *Proc. 2018 IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2018.

[32] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," *arXiv preprint arXiv:1804.02276*, 2018.

[33] ——, "Model-free training of end-to-end communication systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2503–2516, 2019.

[34] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end deep learning of optical fiber communications," *J. Lightw. Technol.*, vol. 36, no. 20, pp. 4843–4855, 2018.

[35] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning," in *Proc. 44th European Conference on Optical Communication (ECOC)*, pp. 1–3, 2018.

[36] Y. Jiang, H. Kim, H. Asnani, S. Kannan, S. Oh, and P. Viswanath, "Turbo autoencoder: Deep learning based channel codes for point-to-point communication channels," in *Advances in Neural Information Processing Systems*, pp. 2758–2768, 2019.

[37] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," *arXiv:1907.08535*, 2019.

[38] T. Matsumine, T. Koike-Akino, and Y. Wang, "Deep learning-based constellation optimization for physical network coding in two-way relay networks," in *Proc. 2019 IEEE International Conference on Communications (ICC)*, 2019.

[39] M. Stark, F. A. Aoudia, and J. Hoydis, "Joint learning of geometric and probabilistic constellation shaping," in *Proc. 2019 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, 2019.

[40] K. Gümüş, A. Alvarado, B. Chen, C. Häger, and E. Agrell, "End-to-end learning of geometrical shaping maximizing generalized mutual information," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2020.

[41] M. Schaedler, S. Calabrò, F. Pittalà, G. Böcherer, M. Kuschnerov, C. Bluemm, and S. Pachnicke, "Neural network assisted geometric shaping for 800Gbit/s and 1Tbit/s optical transmission," in *Proc. Optical Fiber Communications Conference and Exhibition (OFC)*, pp. 1–3, 2020.

[42] E. Balevi and J. G. Andrews, "Autoencoder-based error correction coding for one-bit quantization," *IEEE Trans. Commun.*, 2020.

[43] V. Talreja, T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "End-to-end deep learning for phase noise-robust multi-dimensional geometric shaping," in *Proc. 2020 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2020.

[44] B. Karanov, V. Oliari, M. Chagnon, G. Liga, A. Alvarado, V. Aref, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-end learning in optical fiber communications: Experimental demonstration and future trends," in *Proc. 2020 European Conference on Optical Communication (ECOC)*, pp. 1–4, 2020.

[45] O. Jovanovic, M. P. Yankov, F. Da Ros, and D. Zibar, "Gradient-free training of autoencoders for non-differentiable communication channels," *J. Lightw. Technol.*, 2021.

[46] M. Tanahashi and H. Ochiai, "A multilevel coded modulation approach for hexagonal signal constellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 10, pp. 4993–4997, Oct. 2009.

[47] 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. 25 685–25 699, 2013.

[48] T. Mehmood, M. P. Yankov, S. Iqbal, and S. Forchhammer, "Flexible multilevel coding with concatenated polar-staircase codes for M-QAM," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 728–739, 2020.

[49] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879, 2012.

[50] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, 2015.

[51] C. Stierstorfer and R. F. Fischer, "(Gray) mappings for bit-interleaved coded modulation," in *Proc. 2007 IEEE 65th Vehicular Technology Conference-VTC2007-Spring*, pp. 1703–1707, 2007.

[52] Y. Lee, H. Yoo, J. Jung, J. Jo, and I.-C. Park, "A 2.74-pJ/bit, 17.7-Gb/s iterative concatenated-BCH decoder in 65-nm CMOS for NAND flash memory," *IEEE journal of solid-state circuits*, vol. 48, no. 10, pp. 2531–2540, 2013.

[53] M. Weiner, M. Blagojevic, S. Skotnikov, A. Burg, P. Flatresse, and B. Nikolic, "27.7 A scalable 1.5-to-6Gb/s 6.2-to-38.1 mW LDPC decoder for 60GHz wireless networks in 28nm UTBB FDSOI," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 464–465, 2014.

[54] H. Yoo, Y. Lee, and I.-C. Park, "7.3 Gb/s universal BCH encoder and decoder for SSD controllers," in *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 37–38, 2014.

[55] B. Smith, M. Ardakani, W. Yu, and F. R. Kschischang, "Design of irregular LDPC codes with optimized performance-complexity tradeoff," *IEEE Trans. Commun.*, vol. 58, no. 2, pp. 489–499, 2010.

[56] L. M. Zhang and F. R. Kschischang, "Low-complexity soft-decision concatenated LDGM-staircase FEC for high-bit-rate fiber-optic communication," *J. Lightw. Technol.*, vol. 35, no. 18, pp. 3991–3999, 2017.

[57] T. Koike-Akino, D. S. Millar, K. Parsons, and K. Kojima, "Fiber nonlinearity equalization with multi-label deep learning scalable to high-order DP-QAM," in *Proc. Signal Processing in Photonic Communications*, pp. SpM4G–1, 2018.

[58] R. T. Jones, M. P. Yankov, and D. Zibar, "End-to-end learning for GMI optimized geometric constellation shape," in *Proc. 45th European Conference on Optical Communication (ECOC 2019)*, pp. 1–4, 2019.

[59] T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization to mitigate fiber nonlinearity," in *Proc. 45th European Conference on Optical Communication (ECOC 2019)*, pp. 1–4, 2019.

[60] K. Sugihara, Y. Miyata, T. Sugihara, K. Kubo, H. Yoshida, W. Matsumoto, and T. Mizuochi, "A spatially-coupled type LDPC code with an NCG of 12 dB for optical transmission beyond 100 Gb/s," in *Proc. Optical Fiber Communications Conference (OFC)*, pp. OM2B–4, 2013.

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

[62] ITU-R, "OTU4 long-reach interface," *Report ITU-R G 709.2/T.1131.2*, Jul. 2018.

[63] B. P. Smith, A. Farhood, A. Hunt, F. R. Kschischang, and J. Lodge, "Staircase codes: FEC for 100 Gb/s OTN," *J. Lightw. Technol.*, vol. 30, no. 1, pp. 110–117, 2011.

[64] ETSI TR 102 376-2, "Digital Video Broadcasting (DVB); Implementation guidelines for the second generation system for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: S2 Extensions (DVB-S2X)," Nov. 2015.

[65] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "TensorFlow: A system for large-scale machine learning," in *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pp. 265–283, 2016.

[66] R. T. Jones, T. A. Eriksson, M. P. Yankov, B. J. Puttnam, G. Rademacher, R. S. Luis, and D. Zibar, "Geometric constellation shaping for fiber optic communication systems via end-to-end learning," *arXiv:1810.00774*, 2018.