Rate-Adaptive Concatenated Multi-Level Coding With Novel Probabilistic Amplitude Shaping

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

Published in:
IEEE Transactions on Communications

Link to article, DOI:
10.1109/TCOMM.2022.3162365

Publication date:
2022

Document Version
Peer reviewed version

Rate-Adaptive Concatenated Multi-Level Coding with Novel Probabilistic Amplitude Shaping

Toshiki Matsumine, *Member, IEEE,* Metodi Plamenov Yankov, *Member, IEEE,*
Tayyab Mehmood, *Student Member, IEEE,* and Søren Forchhammer *Member, IEEE,*

***Abstract*—This paper proposes a new probabilistic amplitude shaping (PAS) approach for concatenated two-level multi-level coding (MLC). The proposed system is based on a concatenated forward error correction (FEC) scheme where outer codes are serially concatenated with inner two-level MLC. This concatenated two-level MLC scheme has recently been shown to have a potential for achieving better performance-complexity trade-offs than the conventional bit-interleaved coded modulation (BICM). Meanwhile, PAS has recently been demonstrated to offer remarkable performance gains as well as rate adaptivity. However, the majority of existing works on PAS assume the use of the binary reflected Gray code as a bit-labeling, and its application to coded modulation schemes with other bit-labelings, such as two-level MLC, may not be straightforward. In this paper, we devise a bit-labeling scheme and propose a new PAS structure for an efficient integration of PAS with two-level MLC systems. More specifically, we propose to generate *signed* amplitude symbols with the distribution matcher (DM) for maximizing both coding and shaping gains achieved by two-level MLC and PAS, respectively, while the conventional PAS generates *unsigned* amplitude symbols. It is demonstrated by simulation results that, with 256QAM and inner polar codes, the proposed two-level MLC with PAS simultaneously offers 75% reduction in the number of required inner encoding and soft-decision (SD) decoding operations for given outer and inner FEC code lengths, and up to 0.3 dB performance gain over the conventional PAS scheme.**

***Index Terms*—Multi-level coded modulation, concatenated codes, polar codes, probabilistic amplitude shaping.**

I. Introduction

Concatenated forward error correction (FEC) combined with high-order modulation is an efficient approach for communication systems such as satellite communications [1] and optical communications [2], where not only a high information rate but also high reliability are required. In particular, a serial concatenation of outer and inner codes with hard-decision (HD) and soft-decision (SD) decodings, respectively, has been widely adopted in such systems. For such a scheme, design of inner codes and its combination with high-order modulation such as quadrature amplitude modulation (QAM) plays a key role, since an inner SD decoder is in general more power-consuming than a outer HD decoder [3]–[5]. Considering the practical constraint of power-consumption, development of a capacity-approaching yet low-complexity inner FEC scheme is of significant importance.

Bit-interleaved coded modulation (BICM) is a popular approach to coded modulation in practice where a powerful binary FEC code is combined with QAM in conjunction with bit-wise interleaving. With capacity-approaching codes such as low-density parity check (LDPC) codes and polar codes, BICM systems can approach the constellation constrained capacity of QAM signaling. On the other hand, two-level multi-level coding (MLC) has been recently revived as an inner FEC scheme of a concatenated FEC, and demonstrated to have a potential to achieve better performance-complexity trade-offs than the conventional BICM [6]–[10]. Unlike standard MLC [11], [12], two-level MLC protects only the least reliable bit (LRB) by an inner FEC code, and thus complexity as well as latency associated with SD decoding can be greatly reduced without sacrificing error rate performances.

Constellation shaping is a technique to enhance the achievable information rate (AIR) by imitating the capacity-achieving signal distribution. In this paper, we focus on probabilistic shaping (PS) that manipulates the distribution of equispaced symbols [13], [14], rather than geometric shaping (GS) that changes the locations of equiprobable symbols [15], [16]. In particular, among a number of implementations of PS, we consider the probabilistic amplitude shaping (PAS) structure [17] in conjunction with a constant composition distribution matcher (CCDM) [18] that has gained much attention as it has been shown to be asymptotically optimal in the sense that the normalized Kullback–Leibler (KL) divergence between a target and actual signal distributions vanishes as symbol length increases. In [17], it has been demonstrated that the PAS structure enables us to practically approach the channel capacity using off-the-shelf FEC codes, where a distribution matcher (DM) and the redundancy of a systematic FEC code generate the amplitudes and the signs of pulse-amplitude modulation (PAM) symbols, respectively.

Since its invention, PAS has been extensively studied in the literature not only for additive white Gaussian noise (AWGN) channels but also nonlinear fiber-optic and wireless fading channels [19]–[25], where its remarkable performance improvement as well as rate adaptability have been demonstrated. However, the conventional PAS structure in [17] assumes the use of a binary reflected Gray code (BRGC) [26] as a bit-labeling scheme which is commonly used for BICM systems.

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

Tayyab Mehmood is with Zeuxion ApS, Værløse, Denmark, email: tm@zeuxion.com

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

On the other hand, the focus of this paper is on a two-level MLC scheme, for which the use of a BRGC labeling may result in serious performance degradation. Therefore, the integration of PAS with a two-level MLC system is not straightforward, despite its potential ability to achieve excellent performances.

Recently, two-level MLC schemes combined with PAS have been studied in [27]–[29]. In [28], it has been demonstrated that their MLC with PAS (referred to as "partial-MLC" in [28]) achieves approximately 38–39% reduction in the estimated power consumption compared to the conventional PAS scheme [17]. However, since the methods presented in [27]–[29] are based on the conventional PAS structure, their MLC fail to achieve the target signal distribution (see Fig. 1 in [28]), and thus degrade a shaping gain. More specifically, in [27]–[29], a BRGC labeling was newly introduced in addition to the original bit-labeling designed for two-level MLC [6], [8], [9], in order to compensate for the shaping performance loss. However, since the approaches [27]–[29] have not resolved the design problem of PAS for two-level MLC systems at a fundamental level, their approaches still suffer from the shaping performance loss compared to the conventional PAS [17].

In this paper, we propose a new PAS structure for two-level MLC that is based solely on a single bit-labeling scheme and enables for achieving the same signal distribution as that for the conventional coded modulation with PAS [17]. More specifically, we devise a bit-labeling scheme and propose a new approach to combining a DM and a systematic FEC code for two-level MLC in order to simultaneously maximize coding and shaping gains achieved by two-level MLC and PAS, respectively. The proposed system is evaluated in terms of an AIR, based on which we optimize the balance of coding and shaping rates of the proposed system. From simulation results and the discussion of demodulation and inner SD decoding complexities, we demonstrate that the proposed MLC with PAS achieves a remarkable decoding complexity reduction in terms of the number of required SD decodings, as well as a performance gain over the conventional PAS scheme. Furthermore, we demonstrate by simulations that the proposed approach achieves a better error rate performance than the similar system in [29].

In summary, the main contributions of this paper are as follows:

- We propose a new PAS structure for concatenated two-level MLC based on a novel bit-labeling scheme that enables achieving arbitrary symmetric symbol distribution;
- We perform optimization of DM and FEC code rates based on an approximated AIR;
- We discuss SD decoding complexity in terms of the number of inner SD decodings where significantly lower complexity of the proposed system is demonstrated in comparison with conventional PAS systems;
- We show by simulation results that the proposed two-level MLC with PAS offers significant reduction in SD decoding complexity with similar performance compared to the conventional PAS scheme.
- We also demonstrate the superiority of the proposed system to the similar two-level MLC system with PAS [29] in terms of performance-complexity trade-offs.

The rest of this paper is organized as follows: Section II introduces the background of concatenated two-level MLC. Subsequently, the proposed system model and the proposed PAS approach are described in Section III. Section IV focuses on design of the proposed MLC system in terms of an AIR. In Section V, inner SD decoding complexity of the proposed system is evaluated in comparison with the conventional PAS. Section VI shows the simulation results of the proposed system in terms of an error rate performance, where a significant gain of the proposed scheme in terms of performance-complexity trade-offs is demonstrated. Finally, concluding remarks are given in Section VII.

II. Introduction of Concatenated Two-Level Multi-Level Coding (MLC)

In this section, we briefly introduce the background and the concept of a concatenated two-level MLC scheme that we consider throughout this paper.

A. Outer Codes with HD Decoding

The role of outer codes in a concatenated FEC scheme is to ensure that a strict requirement for high reliability in a communication system is satisfied. The reliability requirement may depend on the specific system, e.g., the target bit error rate (BER) is often specified as 10^{-15} in optical communications. Algebraic codes in conjunction with low-complexity HD decoding, e.g., a single Bose-Chaudhuri-Hocquenghem (BCH) code [30], [31], turbo product codes [32], and staircase codes [33] with BCH component codes would be especially suitable for this purpose.

B. Inner Two-Level MLC

Unlike outer codes, the role of inner codes is to reduce the BER before outer HD decoding such that a HD decoder can bring it down to the target BER. The required maximum BER for a HD decoder to achieve the target BER is referred to as *outer threshold*. Assuming that the signal-to-noise ratio (SNR) is the measure of channel quality, the design objective of inner codes is to achieve this threshold at the smallest possible SNR. Most of modern communication standards adopt BICM [34], [35] combined with capacity-approaching LDPC codes.

The concept of MLC has been originally proposed in [11], and analyzed in [12]. Multi-stage decoding (MSD) is the optimal decoding strategy for MLC in terms of the AIR where each bit level is decoded successively from the LRB to the most reliable bit (MRB) conditioned on a previous decoder estimate. It is also possible for MLC to decode each bit level independently to reduce the decoding latency of MSD [36], [37], at the cost of performance degradation in terms of the AIR due to the suboptimal decoding based on a bit-wise metric, similar to BICM.

A two-level MLC is a special instance of MLC that has been recently studied as an inner code of concatenated FEC,

where only the least reliable level of a modulator input is protected by an SD FEC [6]–[9]. One of the major advantages of two-level MLC over standard MLC is flexibility in its rate design, since the transmission rate of two-level MLC can be adjusted by changing the rate of a single FEC code, similar to BICM. Another advantage of two-level MLC over BICM is that one can reduce the number of SD decodings, or an SD FEC code length, which results in significantly lower decoding complexity as we will discuss in Section V. Although two-level MLC exhibits an error floor since most of modulator input bits remain uncoded, this would not be problematic as long as an error floor appears below an outer FEC threshold. On the other hand, it is difficult to design lower information rates for two-level MLC that encodes only the LRB. However, this issue may be overcome by integrating with PAS in the proposed system.

III. The Proposed System

In this section, we introduce the proposed concatenated two-level MLC scheme and describe how to incorporate PAS with the proposed system. The proposed PAS approach offers an additional rate-adaptivity as well as performance gain for a two-level MLC scheme.

A. Overall System Structure

The transmitter and receiver of the proposed system are depicted in Fig. 1, where we assume 2^m-ary PAM as our modulation format. Let $R_{\text{DM}} \leq m-1$ and N denote a DM rate, i.e., the number of information bits per DM output symbol, and a transmit symbol length, respectively. At the transmitter, the DM first transforms a bit sequence of length $R_{\text{DM}}N$ into an amplitude symbol sequence of length N with a desired PMF. The amplitude symbol sequence is then converted to $(m-1)\times N$ bits for binary FEC encoding based on the specific bit-labeling we assume. Note that $R_{\text{DM}} = m-1$ holds when the amplitude symbols follow uniform distribution. In this work, we consider a CCDM based on arithmetic coding [18] as our DM. However, any other DM implementation, e.g., [25], [38]–[41], is also applicable to the proposed system.

The outer code takes these (shaped) amplitude bits as its input and generates β_{outer} parity bits. The outer code has to be systematic in order to preserve the distribution of amplitude symbols generated by the DM. The outer codeword is then input to the interleaver, where symbol-wise (a symbol consists of $(m-1)$ bits that form an amplitude symbol) interleaving is applied to systematic bits corresponding to the amplitude in order to maintain the amplitude distribution, whereas bit-wise interleaving is applicable to the parity bits corresponding to the sign bits. Also, this interleaver should be sufficiently longer than the symbol length N to average out the BERs of multiple two-level MLC blocks at the receiver side. Afterwards, only the LRB is encoded by a systematic inner SD FEC code, and the other $(m-1)\times N$ bits are directly input to the modulator block without being fed into the inner code. The β_{inner} parity bits of the inner FEC code are then allocated to the LRB in conjunction with β_{outer} parity bits of the outer FEC code[1]. The transmission information rate of this system is given by $R = R_{\text{DM}}$ bits per dimension. Note that, for a given target rate and an outer code rate, the inner code rate of the proposed two-level MLC is in general lower than that of the conventional system. Also, for the same symbol length N, the inner code length is N for the proposed system, whereas that for the conventional PAS scheme is mN. In other words, for a fixed inner code length, the symbol length of the proposed system is m times longer than than of the conventional scheme.

Similar to the conventional PAS [17], it is also possible to extend the proposed system to an even higher rate, i.e., $R > (m-1)$, by allocating additional information bits to the LRB as indicated with a dashed line in Fig. 1. Let γ denote the ratio of these additional information bits to the transmit symbol length N. The overall information rate is then given by

$$R = R_{\text{DM}} + \gamma \text{ [bits/dimension]}. \tag{1}$$

As shown in (1), the proposed system provides rate-adaptivity with two degrees of freedom by adjusting either or both of R_{DM} and γ. We note that the parameter γ as well as inner and outer code rates should be chosen such that the following relationship is satisfied,

$$\gamma N + \beta_{\text{outer}} + \beta_{\text{inner}} = N. \tag{2}$$

To give intuition into how outer and inner code rates, i.e., R_{outer} and R_{inner}, are related to other parameters m and γ, it may be useful to express the inner code rate as

$$R_{\text{inner}} = \frac{\gamma N + \beta_{\text{outer}}}{\gamma N + \beta_{\text{outer}} + \beta_{\text{inner}}} = \frac{\gamma N + \beta_{\text{outer}}}{N}. \tag{3}$$

Meanwhile, the parity bit length of the outer code is given by

$$\beta_{\text{outer}} = \frac{N(m-1+\gamma)(1-R_{\text{outer}})}{R_{\text{outer}}}. \tag{4}$$

From (3) and (4), we see that the parameters R_{outer}, R_{inner}, m and γ satisfy the following relationship,

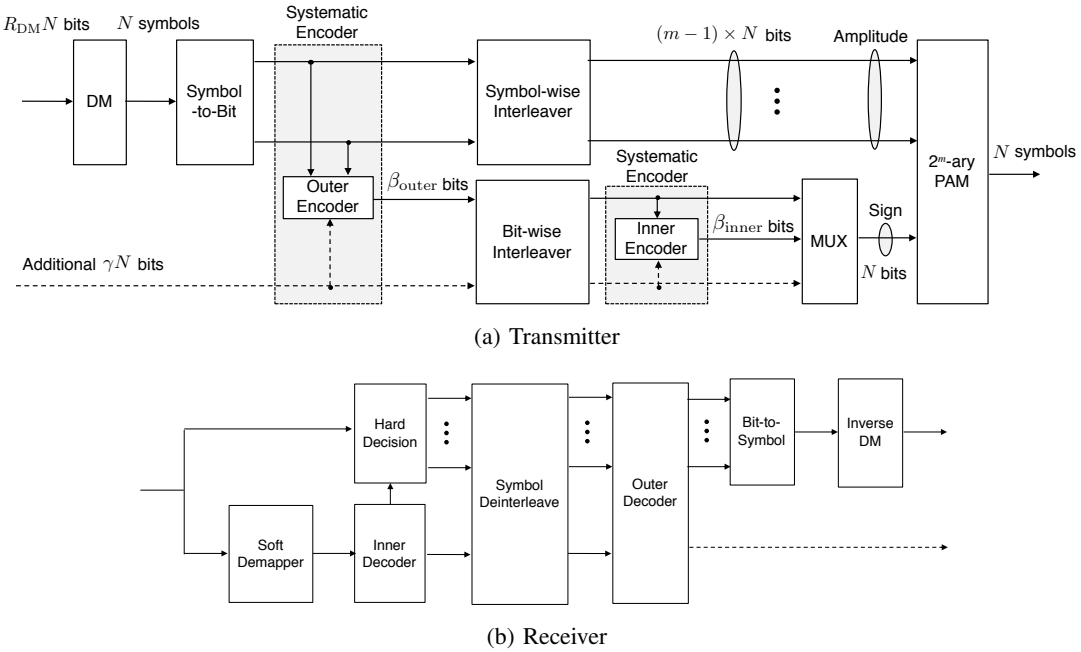
$$(m-1+\gamma)(1-R_{\text{outer}}) + R_{\text{outer}}\gamma = R_{\text{outer}}R_{\text{inner}}. \tag{5}$$

This relationship will be used to design FEC code rates in the proposed system as mentioned in Section VI.

At the receiver side, MSD is employed for decoding MLC where we first decode the LRB by an inner SD decoder and then make a hard-decision on the other bit levels. These bits are input to the symbol deinterleaver to ensure that the averaged BER after deinterleaving is always below the outer threshold. Subsequently, outer HD decoding and an inverse DM operation are performed to retrieve original information bits.

[1]We don't explicitly use interleaver between an inner code and a symbol mapper since we assume memoryless AWGN channels in this paper. However, we may employ symbol interleaver for channels with memory, e.g., wireless fading channels.

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/TCOMM.2022.3162365, IEEE Transactions on Communications



(a) Transmitter

(b) Receiver

Fig. 1: The proposed concatenated two-level MLC with PAS.

B. The Conventional PAS Structure

In what follows, we elaborate more on the details of a PAS structure. We first introduce the conventional coded modulation scheme combined with PAS [17]. Letting $\mathcal{P}_M = \{\pm 1, \pm 3, \ldots, \pm(M-1)\}$ denote a set of $M(=2^m)$-ary PAM constellation points, a transmit symbol $X \in \mathcal{P}_M$ can be decomposed into the amplitude and the sign as $X = A \cdot S$, where the amplitude $A \in \mathcal{A}_{\mathrm{Conv}} = \{1, 3, \ldots, M-1\}$ is chosen from a set of amplitude symbols of PAM $\mathcal{A}_{\mathrm{Conv}}$ whose cardinality is $|\mathcal{A}_{\mathrm{Conv}}| = M/2$, and the sign is $S \in \{-1, 1\}$. In the conventional PAS [17], the DM generates *unsigned* amplitude symbols with desired distribution, and the subsequent systematic FEC encoder generates uniform parity bits which determine the signs of transmit symbols. This PAS structure assumes the property of a BRGC that the MRB affects only signs and does not change amplitudes. Also, the target symbol distribution is assumed to be symmetric with respect to the origin.

In the framework of PAS, Maxwell-Boltzmann (MB) distribution is often adopted as the target distribution since it is a good approximation to the capacity-achieving distribution for AWGN channels [42]. The probability mass function (PMF) of MB distribution is expressed as

$$P_X^{\star}(x) = \frac{\exp(-\lambda x^2)}{\sum_{x \in \mathcal{P}_M} \exp(-\lambda x^2)}, \tag{6}$$

where λ is a parameter to optimize.

As mentioned above, the conventional PAS assumes the symmetric property of a BRGC labeling. However, since this property does not necessarily hold for other bit-labelings, the direct application of the conventional PAS to an MLC-MSD scheme [27], [29] may result in suboptimal shaping performance as discussed in Section VI-B. In order to compensate for this performance loss, the time-variant modulation scheme has been proposed in [27], [29] where two different bit-labeling schemes are combined. More specifically, in addition to the conventional bit-labeling for two-level MLC [6]–[9], BRGC was also introduced for which the conventional PAS can be directly applicable. However, the approaches in [27], [29] have not dealt with the fundamental problem of a PAS design for MLC. On the other hand, we propose a simpler yet more effective solution to the PAS for two-level MLC that is based solely on a single, i.e., time-invariant, bit-labeling.

C. The Proposed PAS for Two-Level MLC

In order for the proposed two-level MLC to maximize both coding and shaping gains simultaneously, the appropriate design of a PAS structure and a bit-labeling plays a key role. More specifically, the proposed bit-labeling scheme is constructed based on the following three design rules;

1) The LRB changes between adjacent symbols,
2) The other levels than the LRB are Gray-labeled,
3) The LRB affects only the signs of PAM symbols.

The first and second conditions are also suggested in [6], [8], [9] for standard two-level MLC without shaping. The first condition is necessary for maximizing coding gain of MLC-MSD schemes. More specifically, this bit-labeling rule ensures that the minimum Euclidean distance increases after the decoding of the LRB in MSD, and thus the uncoded bit levels may be recovered reliably even without FEC. The second condition is for minimizing the BER of uncoded bits after hard decision and thus minimizing the BER at the input of an outer HD decoder. In addition to the above two conditions, we also put a constraint on the bit-labeling that the LRB affects only the sign of PAM symbols for maximizing the shaping gain of PAS. With this proposed bit-labeling, the PAM symbol

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/TCOMM.2022.3162365, IEEE Transactions on Communications

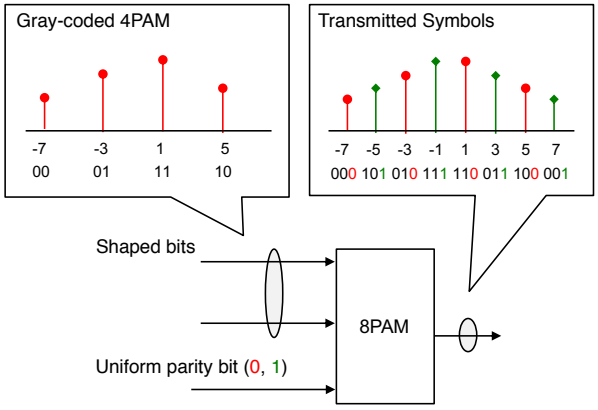


Fig. 2: Example of the proposed PAS with 8PAM.

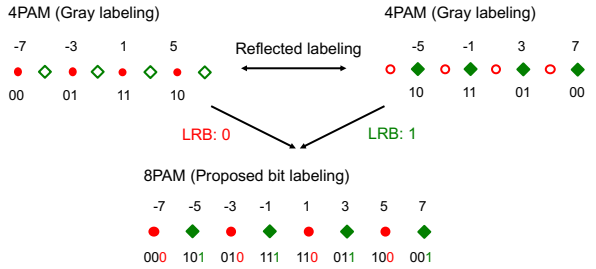


Fig. 3: Construction of the proposed bit labeling for 8PAM.

can be decoupled into amplitude and sign components, and thus the proposed PAS is capable of achieving the same signal distribution as [17].

An example of the proposed PAS with 8PAM is illustrated in Fig 2, where the DM generates amplitudes $\mathcal{A}_{\text{Prop}} = \{-7, -3, 1, 5\}$ with desired distribution, instead of $\mathcal{A}_{\text{Conv}} = \{1, 3, 5, 7\}$. These amplitude symbols are labeled with Gray labeling such that the BER after hard-decision is minimized. Then the uniform parity bits of an inner FEC code are allocated to the LRB that only flips the signs of DM output symbols and does not affect amplitude distribution.

Note that one can systematically construct the proposed bit-labeling for PAM of any size. More specifically, for M-PAM, we start from two Gray-labeled $M/2$-ary PAMs one of which is reflected, and add the additional bit label of 0 and 1 as the LRB to these PAMs, respectively. Then we merge the elements of two $M/2$-ary PAMs alternately for constructing the proposed labeling for M-ary PAM. An example with 8PAM is shown in Fig. 3, where the proposed bit labeling for 8PAM is constructed by combining two Gray-labeled 4PAMs whose bit-labelings are reflected, with the additional LRB of 0 and 1, respectively.

We also note that the proposed system can be extended to two-dimensional QAM in a straightforward manner by allocating two consecutive PAM symbols to real and imaginary components of QAM. Therefore, the resulting PMF of QAM is the product of one-dimensional PMFs of PAM [19].

IV. System Optimization Based on Achievable Information Rate

In this section, we perform optimization of the proposed system in terms of a balance of coding and shaping rates based on AIR. We note here that our objective in this section is to find an (approximated) AIR as a performance metric to be used in system optimization, rather than the exact AIR. We denote probability density function (PDF) of a continuous random variable A as $f_A(a)$, and probability mass function (PMF) of a discrete random variable B as $P_B(b)$. In this paper, we consider a discrete AWGN channel. For simplicity of notation, it is assumed that a channel input symbol X chosen from a set of one-dimensional PAM points $\mathcal{P}_M$, i.e., $X \in \mathcal{P}_M$, and its output is expressed as $Y = X + Z \in \mathbb{R}$, where $Z \in \mathbb{R}$ is a Gaussian random variable with zero-mean and variance σ^2. Letting $P_{\text{av}} = \sum_{x \in \mathcal{P}_M} P_X(x) x^2$ denote the average symbol power, the SNR (in dB) is defined as $\text{SNR} = 10 \log_{10} P_{\text{av}}/\sigma^2$. For two-dimensional QAM signaling, the SNR is given by $\text{SNR} = 10 \log_{10} P_{\text{av}}/2\sigma^2$, which corresponds to two independent PAM signaling.

An achievable transmission rate of PAS is expressed as [43]

$$R_{\text{PAS}} = \left[\mathbb{H}(X) - \mathbb{E}\left[-\log_2 \frac{q(X,Y)}{\sum_{a \in \mathcal{X}} q(a,Y)} \right] \right]^+, \quad (7)$$

where $q(X,Y)$ is a decoding metric. The term $\left[-\log_2 \frac{q(X,Y)}{\sum_{a \in \mathcal{X}} q(a,Y)} \right]$ is called *uncertainty* in [43] and plays an important role in achievable rate calculations. Note that while the term $\mathbb{H}(X)$ depends only on a transmitter, the uncertainty depends on a decoding scheme.

A. Bit-Metric Decoding (BMD)

Let $\mathbf{B} = (B_0, B_1, \ldots, B_{m-1})$ denote a vector of mapper inputs that is mapped onto a signal point $\mathcal{P}_M$ (B_0 is the MRB and B_{m-1} is the LRB). In BMD, decoding is performed based on a bit-wise log likelihood ratio (LLR), which is calculated as

$$\begin{aligned} \ell(B_k) &\triangleq \log \frac{P_{B_k|Y}(B_k = 0|y)}{P_{B_k|Y}(B_k = 1|y)} \\ &= \log \frac{\sum_{x \in \mathcal{P}^0_{M,k}} P_X(x) f_{Y|X}(y|x)}{\sum_{x \in \mathcal{P}^1_{M,k}} P_X(x) f_{Y|X}(y|x)} \end{aligned} \quad (8)$$

where $\mathcal{P}^b_{M,k}$ is the subset of constellation points $\mathcal{P}_M$ whose kth bit label is $b \in \{0, 1\}$.

For PAS with BMD, a DM imposes correlation among modulator input bits $\mathbf{B}$ that are mapped onto a single modulation symbol, and thus the assumption that elements in $\mathbf{B}$ are independent does not hold anymore. In this case, an AIR is given by [43],

$$R_{\text{BMD}} = \left[\mathbb{H}(\mathbf{B}) - \sum_{k=0}^{m-1} \mathbb{H}(B_k|Y) \right]^+, \quad (9)$$

where $[\cdot]^+ = \max(0, \cdot)$. We note that the expression (9) does not take into account the use of outer codes with HD decoding. For a concatenated FEC scheme where HD decoding is employed for the outer code, the overall BER will be minimized

by minimizing the BER after inner decoding, assuming a sufficiently long random interleaver relative to a block length between outer and inner decoders. Since maximization of (9) will minimize the error rate after inner bit-metric decoding for the conventional PAS, we use the expression (9) as a metric for optimization of system parameters.

B. Two-Level MLC with MSD

Let ϵ_{MRB} and ϵ_{LRB} denote the BERs of the MRBs after hard decision and LRB after inner SD decoding. Regarding the system blocks inside the outer encoder and the outer decoder in Fig. 1 as a binary-output channel, the proposed system can be seen as the PAS system with HD decoding [43], [44] where the average BER before HD decoding is given by $\bar{\epsilon} = \left((m-1)\epsilon_{\text{MRB}} + \epsilon_{\text{LRB}}\right)/m$. For this system, an AIR is given as

$$\begin{aligned} R_{\text{TL-MLC}} &= \left[\mathbb{H}(\mathbf{B}) - m\mathbb{H}_{\text{b}}\left(\bar{\epsilon}\right)\right]^{+} \\ &= \left[\mathbb{H}(B_0,\ldots,B_{m-2}|B_{m-1}) + \mathbb{H}(B_{m-1}) - m\mathbb{H}_{\text{b}}\left(\bar{\epsilon}\right)\right]^{+} \end{aligned} \tag{10}$$

where $\mathbb{H}_{\text{b}}(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy function.

For the proposed scheme, the rate of the LRB is reduced by the inner code at the transmitter. Assuming that the input to the inner coder follows a Bernoulli distribution with parameter $1/2$, each codeword of the inner code appears with equal probability $1/2^k$. Also, we assume a random interleaver after inner encoding (for analytical purpose), such that the interleaved LRB bits follow an independent and identically distributed (i.i.d.) distribution. Therefore the entropy of the LRB is $\mathbb{H}(B_{m-1}) = R_{\text{inner}}$, and thus we have

$$R_{\text{TL-MLC}} = \left[\mathbb{H}(B_0,\ldots,B_{m-2}|B_{m-1}) + R_{\text{inner}} - m\mathbb{H}_{\text{b}}\left(\bar{\epsilon}\right)\right]^{+} \tag{11}$$

The above expression may not be convenient as a performance metric to be used in system optimization in terms of a balance of coding and shaping rates since it depends not only on a shaping parameter (symbol distribution) but also a specific inner code. Therefore, we consider the ideal case where the inner code is capacity-achieving as we are interested in the ideally achievable performance.

Assuming ideal capacity-achieving inner codes, the BER of the LRB after inner SD decoding is zero at a target information (code) rate. In this case, the outer decoder only takes erroneous bits from the MRBs, and the receiver and the expression in (10) can be separated into two parts depending on different decoding schemes, i.e., bit-metric decoding for the LRB, and binary hard-decision decoding for the MRBs based on the correctly decoded LRB where the decoder input BER is ϵ_{MRBs}. These two decoding schemes result in different receiver uncertainties [43]. More specifically, the rate of the LRB can be expressed by the mutual information, whereas the rate for the remaining $(m-1)$ MRBs is given by an AIR for binary HD decoding. Therefore the rate for this system is given by

$$R^{\star}_{\text{TL-MLC}} = \mathbb{H}(\mathbf{B}) - \underbrace{(m-1)\mathbb{H}_{\text{b}}(\epsilon_{\text{MRBs}})}_{\text{uncertainty of the MRBs}} - \underbrace{\mathbb{H}(B_{m-1}|Y)}_{\text{uncertainty of the LRB}}. \tag{12}$$

The expression (12) generally serves as an upper bound of $R_{\text{TL-MLC}}$ (11) as (12) assumes the ideal case where the inner decoder outputs no bit errors. In fact, the existence of the outer code at the LRB does not increase an AIR due to the data processing inequality. Unlike (10), the expression (12) does not depend on performance of a specific inner code, and thus we may use (12) as a performance metric in the optimization of the coding and shaping rates.

C. Balancing Coding and Shaping Rates

A target information rate of the proposed system, given by (1), is determined by a DM rate R_{DM} and an additional information rate γ. Note that from the condition (2), the number of FEC parity bits corresponds one-to-one with an additional rate γ. In order to achieve good error rate performances, a proper rate allocation to coding and shaping is essential. In what follows, we describe how to find the optimal balance of coding and shaping redundancies.

Assuming the CCDM with infinite symbol length, the gap between an actual DM rate and an entropy $\mathbb{H}(A)$ vanishes, i.e., $R_{\text{DM}} = \mathbb{H}(A)$. Also, assuming that the sign bits are uniformly distributed, $\mathbb{H}(X)$ is given as $\mathbb{H}(X) = \mathbb{H}(A) + 1$. Therefore, once we get the optimal entropy $\mathbb{H}(X)$ for a given target information rate R_{target}, we may find the optimal parameter γ as,

$$\gamma = R_{\text{target}} - \underbrace{\left(\mathbb{H}(X) - 1\right)}_{=\mathbb{H}(A)}. \tag{13}$$

Let $\text{SNR}_{\text{TL-MLC}}(R_{\text{target}}, \lambda)$ denote the SNR (in dB) where an AIR of two-level MLC with a MB parameter λ reaches a target rate R_{target}. Similarly, we denote a required SNR (in dB) for the optimal continuous Gaussian signaling to achieve a target rate R_{target} as $\text{SNR}_{\text{capacity}}(R_{\text{target}}) = 10\log_{10}(2^{2R_{\text{target}}} - 1)$. We then define the gap to capacity (in dB) as follows,

$$\Delta\text{SNR}(\lambda) = \text{SNR}_{\text{TL-MLC}}(R_{\text{target}}, \lambda) - \text{SNR}_{\text{capacity}}(R_{\text{target}}). \tag{14}$$

Our objective here is to find the optimal parameter λ, or equivalently, the optimal entropy $\mathbb{H}(X)$ under the assumption of MB distribution, that minimizes the gap to capacity by changing the parameter of the MB distribution λ. The optimal MB distribution parameter $\lambda^{\star}$ is then expressed as,

$$\lambda^{\star} = \arg\min_{\lambda} \Delta\text{SNR}(\lambda) \tag{15}$$

The definition of the gap to capacity is the same for the conventional PAS with BMD where $\text{SNR}_{\text{TL-MLC}}$ is replaced by SNR_{GMI}.

In Fig. 4, the performances of two-level MLC and the conventional coded modulation schemes employing 16PAM in terms of the gap to capacity (14) are shown. Note that $\mathbb{H}(X) = 4$ corresponds to uniform 16PAM without shaping. From this figure, it is observed that there exists an optimal entropy $\mathbb{H}(X)$ for both the proposed two-level MLC and the conventional scheme, that minimizes the gap to capacity depending on the target spectral efficiency. More specifically, for the target spectral efficiency of 3.2 bits per dimension, we observe that with the optimized PMF, the distances from the

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/TCOMM.2022.3162365, IEEE Transactions on Communications

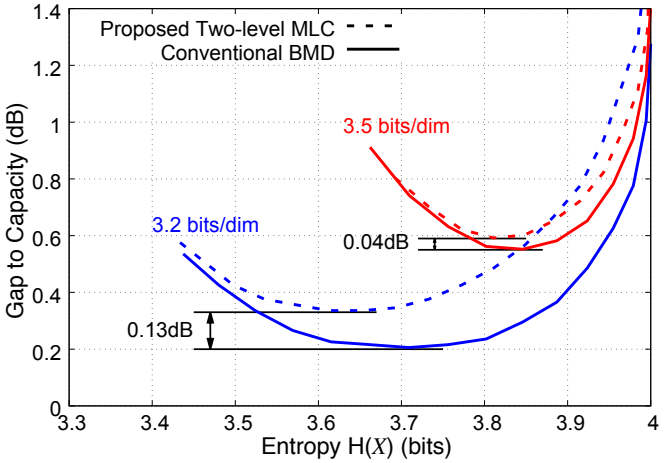


Fig. 4: The gap to capacity performances of the proposed two-level MLC and the conventional BMD with 16PAM for different target rates R_{target}.

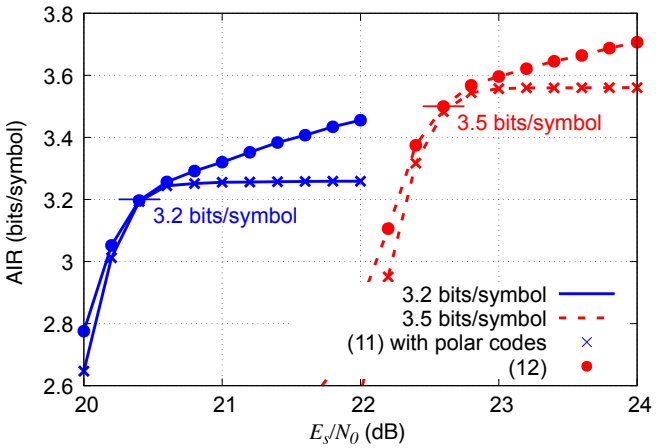


Fig. 5: Comparison of two AIR expressions for the approximated system (12) and the original system (10) with an inner polar code of length 1024 bits for 16PAM.

Shannon limit of AIRs of the proposed two-level MLC and the conventional scheme are about 0.33 dB and 0.2 dB, respectively. Also, we observe that the performance gap between the proposed two-level MLC and the conventional scheme becomes smaller when the target rate is 3.5 bits per dimension. We note that when HD decoding is employed for all bit levels, the gap to capacity is around 2 dB even with the optimized MB distribution [44]. Therefore, we see that the proposed system efficiently reduces the gap to capacity by using a power-consuming SD decoder only at one bit level. In the simulation results of Section VI, DM and polar code rates are optimized based on this analysis.

We next verify the impact of the approximation (12) based on simulations. In Fig. 5, we compare the two AIR expressions (12) and (10) with inner polar codes for target rates of 3.2 and 3.5 bits per symbol with 16PAM. The parameters of MB distribution are optimized based on Fig. 4. It is observed from Fig. 5 that AIR curves with (10) are very close to the approximated ones (12) at their target information rates. Note that (10) is lower than (12) at high SNR regions, due to the redundancy of inner polar codes. Also, we observe a slight gap between the two curves in the case with 3.5 bits per symbol, which stem from a performance loss of practical inner polar codes from the capacity. However, we would like to emphasize that our objective here is to find a metric for system optimization rather than the exact AIR, for which (12) is still useful in the relevant regions.

It should be noted that the AIR represents an upper performance limit of the given system and is achievable with an ideal, infinitely long FEC code. As such, the results in Fig. 4 highlights the asymptotic disadvantage of the proposed concatenated MLC with outer codes with respect to the conventional scheme [17] without outer codes and is thus a relevant figure of merit for communication systems. However, in practice, error performance with outer codes would be of interest for both the proposed and the conventional systems to achieve the strict BER requirement of 10^{-15}. As will be seen in Section VI, when an outer code is employed for both schemes, the proposed scheme can outperform the conventional PAS in terms of an error rate.

We note that another important performance metric in practice would be error exponents [45]–[49], which allows us to analyze how the expected block error rate (BLER) decays for code ensembles such as random code ensembles with its length. For example, error exponents for BICM [46], [48] and MLC with MSD [47] have been studied. Furthermore, error exponents of BICM with probabilistic shaping has been analyzed in [49]. These analyses could be potentially used for analyzing how the proposed scheme behaves with respect to its length. However, we evaluate the performance of the proposed scheme mainly using Monte-Carlo BER simulations in Section VI rather than the error exponent analysis that generally provides the expected BLER of code ensembles instead of BER, since such an analysis may not be straightforward for the proposed concatenated MLC where the overall error probability depends on the BER before outer decoding.

V. Discussion of Relative Decoding Complexity

In this section, we discuss the computational complexity required for the proposed inner two-level MLC in comparison with the conventional PAS scheme [17]. More specifically, we focus on relative inner SD decoding complexity, since the benefit of the proposed scheme in terms of complexity mostly comes from reduction in SD decoding complexity. While demodulation complexity may be also reduced by the proposed two-level MLC, we do not perform an in-depth complexity analysis that includes both decoding and demodulation complexities as it will highly depend on practical implementations.

Again, the key feature of the proposed two-level MLC is that only one bit level is protected by an inner code, and this structure allows us to use the same number of shorter inner codes, or fewer inner codes of the same length for a given outer code, compared to the conventional scheme. In this work, we use the number of required inner SD decoding operations for fixed outer and inner code lengths as our complexity measure.

The relative complexity saving by the proposed MLC is evaluated in terms of reduction in the number of SD decodings [37] compared to the conventional PAS, assuming that two codes of the same length have similar decoding complexities regardless of their rates[2]. This relative complexity reduction is general in the sense that it is independent of the specific inner FEC code.

As mentioned in Section III, for the same inner code length (while their rates are different), the symbol length of the proposed two-level MLC is m times longer than that of the conventional scheme [17]. This means that the proposed MLC requires m times fewer inner encoding and decoding operations compared to the conventional scheme in order to transmit the same number of information bits at the same information rate. Therefore, we conclude that SD encoding and decoding complexities of the proposed MLC are approximately $1/m$ of those for the conventional scheme. The effect of this complexity saving becomes more significant for a higher order modulation. For instance, for 16PAM (256QAM), the number of required inner SD encodings and decodings for the proposed MLC scheme is only 25% of that for the conventional scheme.

VI. Simulation Results

In what follows, we evaluate the performance of the proposed MLC with PAS in comparison with the conventional PAS, and also the similar MLC scheme with PAS [29] over AWGN channels. We make performance comparisons with the same inner code length, which results in significant reduction in terms of inner SD encodings and decodings of the proposed two-level MLC compared to the conventional schemes, as discussed in Section V.

As our outer code, without loss of generality, we assume the staircase code defined in the ITU-T G.709.2 recommendation with code rate $R_{\text{outer}} = 239/255 \approx 0.937$, and the (30832, 30592) BCH code of rate $R_{\text{outer}} = 30592/30832 \approx 0.992$. The BER thresholds of the staircase and the BCH codes for the target BER of 10^{-15} are 5.0×10^{-3} [2], and 5×10^{-5} [30], respectively[3].

For inner codes, we consider polar codes [50] due to their flexible rate design without changing encoding and decoding structures, as well as their good performance at medium-to-short code length. More specifically, we assume systematic polar codes [51], [52] as they achieve better BER performance than their non-systematic counterparts. For decoding of polar codes, we use SCL decoding with list size 16 in conjunction with 8-bit CRC [53] in all cases[4]. We use existing polar code designs assuming binary AWGN channels and SC decoding for both the proposed MLC and the conventional systems. More specifically, letting $\text{SNR}_{\text{BIAWGN}}(R)$ denote the minimum required SNR to achieve a rate R suggested by the AIR of a binary AWGN channel, we construct polar codes using the Tal-Vardy's method [54] with the design SNR given by $\text{SNR}_{\text{design}} = \text{SNR}_{\text{BIAWGN}}(R_{\text{inner}})$. We note that while a theoretic approach seems to be challenging, there may be room for performance improvement by designing polar codes for our system based on heuristics [55], [56], which we leave as future work. As a DM, we assume CCDM [18] of length 1024 real symbols in all cases. Nevertheless, we would like to emphasize that any other systematic FEC codes, as well as DM implementations are applicable to the proposed system.

A. Comparison with the Conventional PAS Scheme

In what follows, we compare the BER performances before HD decoding, which is often referred to as "pre-FEC BERs", of the proposed MLC and the conventional scheme [17]. With the assumption of the ideal random interleaving between outer and inner FEC codes, the target BER of 10^{-15} will be achieved when a pre-FEC BER is below an outer BER threshold. The balance of shaping and code rates are optimized based on the analysis in Section IV-C, and the inner code rate is determined to satisfy (5) for a given outer code rate. The code rates for both the proposed two-level MLC and the conventional scheme used in the following simulations are summarized in Table I. The code rates of the conventional scheme are also optimized in terms of the AIR, similar to the proposed system.

1) 16QAM: In Fig 6, we plot the pre-FEC BERs with 16QAM for which the proposed MLC requires 50% fewer SD decodings compared to the conventional scheme as discussed in Section V. Figure 6a shows pre-FEC BER performances at a information rate of 2.8 bits per symbol. Performances with the outer BCH and the staircase codes, indicated by the solid and dashed curves, respectively, target different BER thresholds. The corresponding outer BER thresholds for the staircase code and the BCH code, i.e., 5.0×10^{-3} and 5.0×10^{-5}, respectively, are indicated by horizontal dashed and solid lines, respectively. Also, the minimum required SNR for the proposed MLC to achieve the rate 2.8 bits per symbol suggested by the AIR result in Fig. 4, which we refer to as the "AIR limit", is indicated by a vertical solid line. From this figure, we observe that when the outer code of higher rate, i.e., the BCH code of rate 0.992, is employed, the proposed MLC exhibits an error floor above the corresponding threshold of 5.0×10^{-5} due to uncoded MRBs, while the conventional PAS achieves the outer threshold at an SNR of around 9.6 dB. The reason for this error-floor and a ways to mitigate it is presented later in this Section. On the other hand, when the staircase code of rate 0.937 is employed as an outer code, both the proposed MLC and the conventional scheme reach the corresponding

[2]We note that SD decoding complexity can vary even with the same length depending on other code parameters. For example, for LDPC codes, decoding complexity depends on particular degree distribution. Therefore, in order to evaluate their encoding and decoding complexities more precisely, an LDPC code-specific complexity measure should be introduced as in [7], [9], which is beyond our scope in this paper. On the other hand, our assumption appears to be true for polar codes.

[3]We note that since the inverse DM operation increases the BER even after HD decoding [40], a little smaller BERs than the thresholds may be required before HD decoding in practice for achieving the target overall BER of 10^{-15}.

[4]The use of CRC bits may be not necessary when powerful outer codes are available, since concatenation with CRC codes improves performance of SCL decoding at a low BER region, which may be already below an outer threshold.

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/TCOMM.2022.3162365, IEEE Transactions on Communications

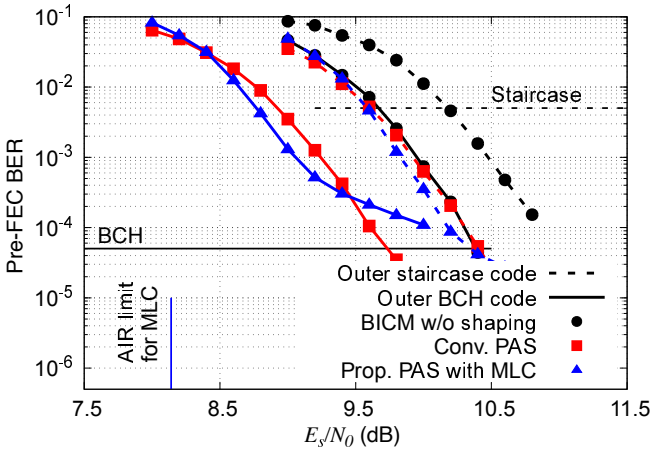
TABLE I: FEC code rates used in the simulation for the proposed MLC and the conventional scheme [17].

Modulation	information rate (bits/symbol)	Inner code rate (Proposed)	Inner code rate (Conventional)	Outer code rate
16QAM	2.8	0.71	0.83	0.937
		0.58	0.78	0.992
	3.2	0.81	0.89	0.937
		0.71	0.84	0.992
256QAM	6.4	0.79	0.93	0.937
		0.58	0.88	0.992
	7.0	0.92	0.98	0.937
		0.71	0.93	0.992

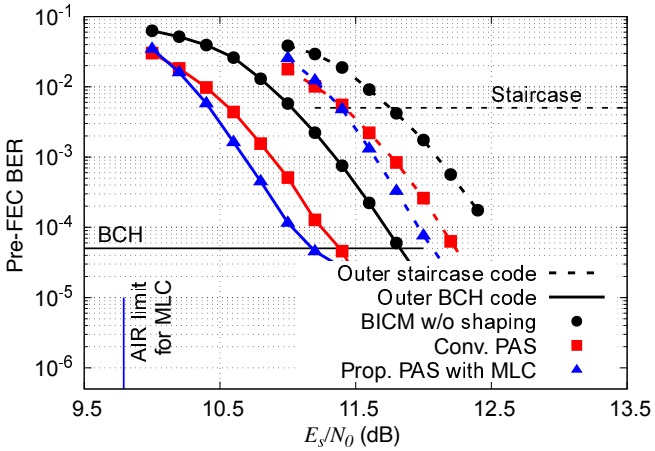
threshold of 5.0×10^{-3} at an SNR of 9.6 dB, which is approximately 1.7 dB away from the AIR limit. Therefore, the proposed MLC with the outer staircase code achieves similar performance to that of the conventional scheme with either the outer BCH code or the staircase code, and the shaping gain over BICM without shaping in this case is about 0.6 dB. We note that while the proposed and the conventional PAS schemes with the outer BCH code achieve similar performance at the outer threshold, there is a BER region where the proposed MLC clearly outperforms the conventional scheme, e.g., around 10^{-3}. This indicates that the proposed scheme may achieve better performance when an outer threshold is fixed at a little higher level than the selected BCH code. In the simplest implementation, this is achieved by choosing an outer code of lower rate, in which case the penalty is a slight degradation of the net data rate.

Figure 6b shows the pre-FEC BERs for a information rate of 3.2 bits per symbol with 16QAM. Similar to Fig. 6a, the outer BER thresholds and the AIR limit of the proposed MLC for this information rate are indicated in the figure. It is observed from this figure that the error floor of the proposed MLC appears at lower BER as the information rate increases, since the uncoded bits become more reliable for higher information rates, i.e., higher SNR regions. Also, with the outer BCH code, the proposed MLC achieves the corresponding BER threshold at an SNR of 11.2 dB, which is about 1.6 dB away from the AIR limit, and 0.1 dB better than the conventional scheme, even with the 50% fewer of SD decodings. On the other hand, when the outer staircase code is employed, the performance gains of both the proposed and conventional schemes over BICM with shaping decrease, since the outer code rate is low for the information rate.

2) 256QAM: Figure 7 presents the pre-FEC performances with 256QAM, where the proposed MLC requires 75% fewer SD decodings compared to the conventional scheme. The outer thresholds and corresponding AIR limits are also indicated as in Fig. 6. Figure 7a shows the pre-FEC BERs with 256QAM at information rates of 6.4 bits per symbol. From this figure, it is observed that, both the proposed MLC and the conventional scheme achieve better performance with the outer BCH code than the case with the outer staircase code. More specifically, with the outer BCH codes, the proposed MLC achieves the corresponding threshold 5×10^{-5} at an SNR of 21.4 dB, similar to the conventional scheme. In this case, the performance gain over BICM without shaping is about 1.0 dB, and the gap fom the AIR limit is about 1.9 dB. Note that the proposed scheme with the outer BCH code clearly outperforms the conventional scheme in a range of 10^{-4}–10^{-3}, while exhibits



(a) 2.8 bits/symbol.



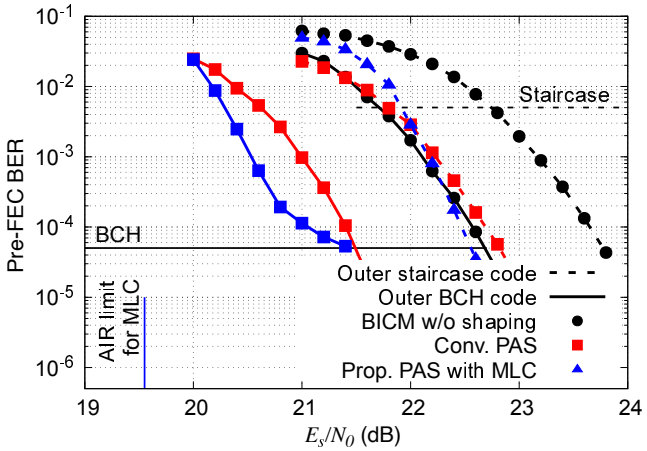
(b) 3.2 bits/symbol.

Fig. 6: Pre-FEC BER comparison with the conventional PAS scheme [17] with 16QAM and polar code length of 1024 bits.

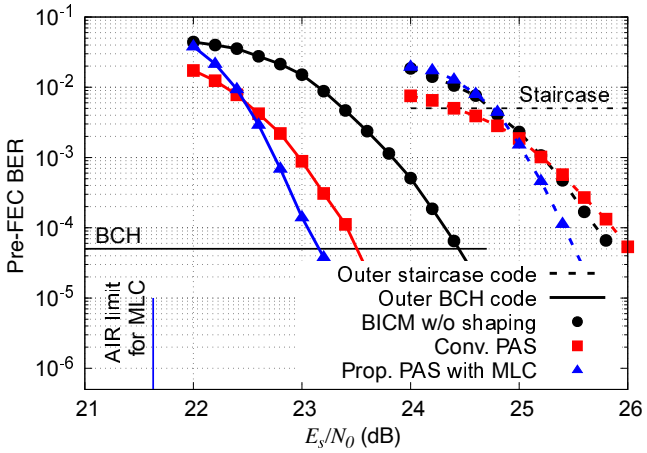
an error floor at around the threshold. The error floor will be mitigated by choosing a lower outer code rate, similar to the case in Fig. 6a.

We also plot the pre-FEC BER with 256QAM for a information rate of 7.0 bits per symbol in Fig. 7b. Similar to Fig. 7a, concatenation with the outer BCH code results in better performance for both the proposed MLC and the conventional scheme compared with the outer staircase code. More specifically, the proposed MLC achieves the threshold of the BCH code at an SNR of 21.8 dB which is about 1.6 dB away from the AIR limit, and 0.3 dB better than the conven-

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/TCOMM.2022.3162365, IEEE Transactions on Communications



(a) 6.4 bits/symbol.



(b) 7.0 bits/symbol.

Fig. 7: Pre-FEC BER comparison with the conventional PAS scheme [17] with 256QAM and polar code length of 1024 bits.

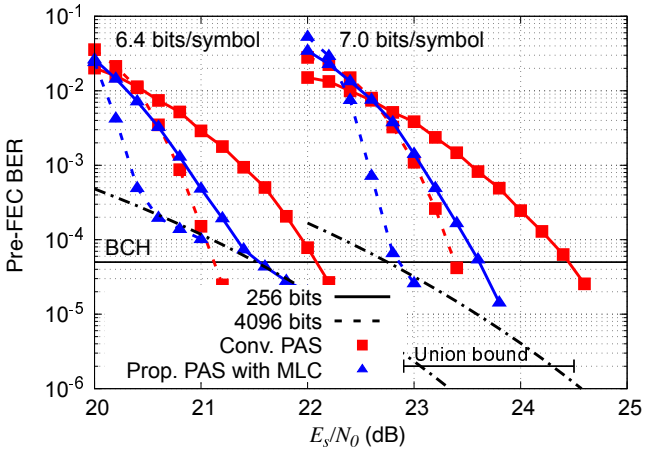


Fig. 8: Pre-FEC BER comparison for different inner code lengths (256, 4096 bits) with 256QAM and the union bounds on the BER of the MRBs.

tional scheme. Also, we observe that when concatenated with an outer code of relatively low rate, both the proposed MLC and the conventional scheme achieve almost no gain over the system without shaping. This suggests that the outer code rate should be sufficiently small, considering the different roles of outer and inner FEC codes. Namely, decreasing the outer code rate to increase the outer threshold degrades error reducing performance of the inner code. Note that the inner code rate of the system with shaping is higher than that for the system without shaping for given outer and target information rates.

3) Impact of Inner Code Length: So far, we have exclusively focused on inner polar codes of length 1024 bits. In what follows, we examine the impact of an inner code length on pre-FEC BER performance. Figure 8 shows pre-FEC BER performances for the same information rates as Fig. 7, but for different inner polar code lengths of 256 and 4096 bits. The outer BCH code of rate $R_{\text{outer}} = 0.992$ is assumed as an outer code. We observe that the error floor of the proposed MLC becomes prominent for a lower information rate, and a longer inner code length. In particular, with the polar code length of 4096 bits, the error floor appears above the BER threshold of the outer BCH code for 6.8 bits per symbol. This indicates that for longer inner codes and lower information rates, lower outer code rates (higher outer thresholds) are required for the proposed two-level MLC. Nevertheless, the proposed two-level MLC still benefits from the longer inner code length when outer threshold is set to be a little higher. Also, for 7.0 bits per symbol, we see that the performance gains of the proposed MLC over the conventional scheme at the outer threshold are about 0.8 dB and 0.6 dB for the inner code lengths of 256 bits and 4096 bits, respectively.

In order to verify these observations from a theoretical perspective, we plotted the union bound on the BER of the MRBs in Fig. 8 (see Appendix for the derivation). The union bound will serve as an upper bound of the pre-FEC BER at a high SNR region where the pre-FEC BER is dominated by the BER of the MRBs. We note that, when the inner code exhibits an error floor, the derived bound will not strictly serve as an upper bound of the pre-FEC BER. However, as long as the error floor is lower than the outer threshold, the derived bound is useful to estimate error floor of the proposed scheme in a BER region of interest. This is exemplified in Fig. 8 using inner polar codes, where it is confirmed that the simulated pre-FEC BER curves agree well with the union bounds around the region of interest of the decoding threshold of the outer code, indicating that the error floor is dominated by the BER of the uncoded MRBs and thus irrespective of the inner code length. This is further supported by the fact that the BER curves of the proposed MLC with the two different inner code lengths converge asymptotically at a high SNR region.

4) Performance with LDPC Codes: As already mentioned, the proposed system does not assume any specific inner code. In what follows, we evaluate performances of the proposed two-level MLC with inner SD LDPC codes, that are widely adopted in the modern communication standards. Specifically we assume LDPC codes of length 64800 bits from the DVB-S2 standard [31] in conjunction with sum-product decoding with a maximum iteration count of 50 for both the proposed and the

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/TCOMM.2022.3162365, IEEE Transactions on Communications

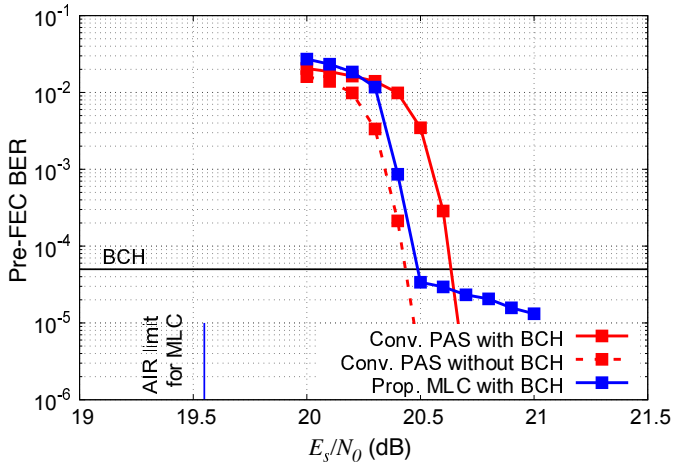


Fig. 9: Pre-FEC BERs with DVB-S2 LDPC codes for a information rate 6.4 bits per symbol with 256QAM.

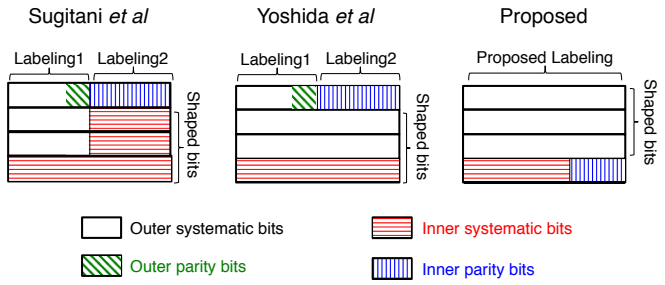


Fig. 10: Frame structures of the proposed system, and the similar approaches proposed by Sugitani *et al.* [27], [28] and Yoshida *et al.* [29] with 16PAM. Corresponding bit-labeling schemes are shown in Table II. Note that outer parity bits are included in inner systematic bits for the proposed system.

conventional schemes. The resulting symbol lengths for two-level MLC and the conventional scheme are therefore 64800 and 16200 symbols, respectively. The inner code rates are set to 0.90 and 0.75 for the conventional scheme and the proposed MLC, respectively[5]. Figure 9 presents a pre-FEC performance comparison of the proposed MLC and the conventional scheme with DVB-S2 LDPC codes for 6.4 bits per symbol. The BCH code with a BER threshold of 5×10^{-5} is assumed as an outer code. It is observed from Fig. 9 that, with the outer BCH code, the proposed two-level MLC achieves better BER performance than the conventional scheme at the threshold. More specifically, the performance gain of the proposed two-level MLC is 0.2 dB. This demonstrates the effectiveness of the proposed scheme regardless of a specific inner FEC code.

From the AIR result in Fig. 4, we observed that the AIR curve of the proposed MLC is slightly inferior to that of the conventional scheme. In order to investigate the operational meaning of Fig. 4, we also plot pre-FEC BER of the conventional scheme with the stand-alone LDPC code, i.e., without the outer BCH code. We see that, the pre-FEC BER of the conventional scheme without the outer BCH code begins to fall at a slightly smaller SNR than the proposed MLC with concatenated BCH and LDPC codes. This performance hierarchy agrees with the AIR result in Fig. 4, indicating that if capacity-achieving codes are available, the conventional scheme with a single code has a potential to outperform the proposed two-level MLC. However, in practice, it would be challenging for a single code to approach the channel capacity under the strict BER requirement as low as 10^{-15}, since capacity-approaching codes such as LDPC codes typically exhibit an error floor. On the other hand, when concatenated with outer codes, performance of the conventional scheme degrades due to the rate loss associated with outer codes, and thus the proposed two-level MLC may achieve better performance as observed in Fig. 9.

[5]For the proposed MLC, the optimized inner code rate based on the AIR analysis is around 0.58, however, we increased it to 0.75 such that the error floor appears below the threshold.

B. Comparison with Similar MLC-PAS Schemes

Finally, we compare the proposed system with similar PAS approaches proposed for two-level MLC in [27]–[29]. Figure 10 illustrates the frame structures of the proposed two-level MLC and the related approaches [27]–[29] employing 16PAM. As shown in Fig. 10, unlike the proposed system, the approaches proposed in [27]–[29] utilize a time-variant modulation scheme, where two different bit-labeling schemes are combined. These bit-labeling schemes are compared in Table II with the proposed bit-labeling scheme. The "Labeling1" is the bit-labeling proposed for two-level MLC without PAS [6]–[9], and the "Labeling2" is the BRGC labeling which is commonly used for the conventional PAS. As shown in Table II, the "Labeling1" cannot be decoupled into amplitude and sign components, whereas this is possible for the proposed bit-labeling as described in Section III. Therefore, the approaches [27]–[29] may fail to achieve the target MB distribution. Note that the MRB determines the signs of modulation symbols for the system [29], while the LRB changes the signs of symbols in the proposed system.

In [27], [28], it has been proposed to employ both conventional PAS scheme [17] and two-level MLC in a single transmission block. More specifically, the "Labeling1" corresponds to two-level MLC, and the "Labeling2", i.e., the BRGC labeling, corresponds to the conventional scheme. However, the use of the conventional scheme increases decoding complexity and latency for a fixed block length, and thus negates the advantage of two-level MLC. On the other hand, the approach in [29] reduces SD decoding complexity by encoding only the MRB and the LRB with an SD FEC code for the "Labeling2". However, the inner FEC code length is still longer than that of ours for a fixed block length. More specifically, letting N and N_{BRGC} denote a block length and the number of symbols with the BRGC labeling ("Labeling2"), respectively, the inner code lengths of the conventional systems in [27], [28] and in [29] are $N+(m-1)N_{\mathrm{BRGC}}$ and $N+N_{\mathrm{BRGC}}$, respectively for 2^m-ary PAM, whereas that of the proposed system is N. The ratio of the BRGC labeling ("Labeling2") to a symbol length N is determined by a code rate of an inner code. Since parity bits of an inner code are allocated to the MRB of the BRGC labeling, a lower code rate of an inner code means a higher

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/TCOMM.2022.3162365, IEEE Transactions on Communications

TABLE II: The proposed bit-labeling and those in the related systems [29] for 16PAM.

X	15	13	11	9	7	5	3	1	-1	-3	-5	-7	-9	-11	-13	-15
Proposed	0000	1001	0010	1011	0110	1111	0100	1101	1100	0101	1110	0111	1010	0011	1000	0001
Labeling1 [29]	0101	0100	0111	0110	0011	0010	0001	0000	1001	1000	1011	1010	1111	1110	1101	1100
Labeling2 [29]	0101	0100	0111	0110	0011	0010	0001	0000	1000	1001	1010	1011	1110	1111	1100	1101

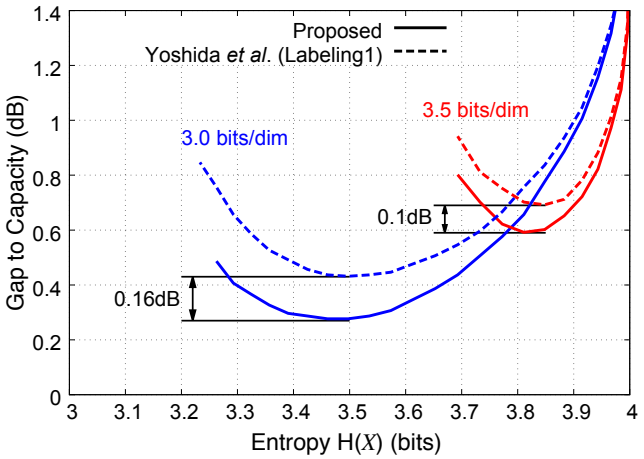


Fig. 11: The gap to capacity performances of the proposed system and the similar system in [29] with 16PAM.

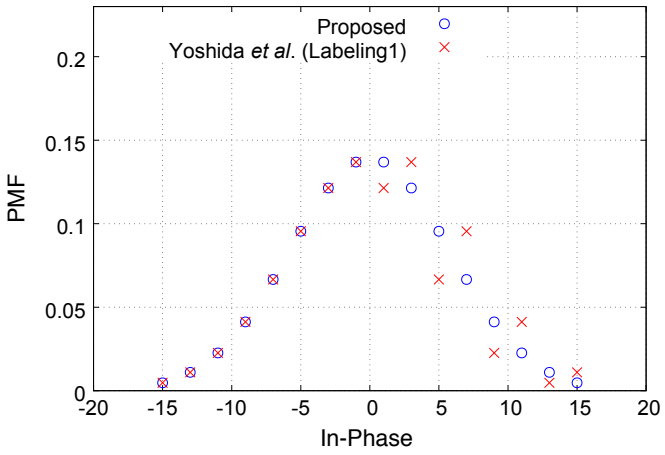


Fig. 12: The PMF of 16PAM with the bit-labeling in Table II [29] for a target rate 3.0 bits per dimension (6.0 bits per complex symbol).

ratio of the BRGC labeling ("Labeling2") to the "Labeling1", and vice versa.

We next compare shaping performances of the proposed system and the similar system [29]. In order to evaluate pure shaping gains, we compare the proposed system with the first time block of the system [29] (corresponding to the "Labeling1") in Fig. 10 for which only the LRB is encoded by an SD FEC code, similar to our system. Since the only difference between the compared two systems is the PAS approach, i.e., the PMF of transmit signals, they result in the same AIR performance if their PMFs are same. Figure 11 shows the gap to capacity performances of the proposed and similar two-level MLC schemes for target information rates 3.0 and 3.5 bits per symbol with 16PAM (equivalent to 6.0 and 7.0 bits per complex symbol with 256QAM), where an AIR is computed based on (12). From Fig. 11, we observe that the proposed scheme offers about 0.16 dB and 0.1 dB gains for 3.0 and 3.5 bits per symbol, respectively. The corresponding optimized PMFs for 3.0 bits per dimension (6.0 bits per complex symbol) are shown in Fig. 12. As mentioned earlier, since the "Labeling1" cannot be decomposed into amplitude and sign components, the target MB distribution may not be achieved. Therefore, the AIR performance gaps in Fig. 11 are interpreted as loss of shaping gain due to the suboptimal shaping of the system [29] as shown in Fig. 12.

Figure 13 shows pre-FEC BER performance comparisons of the proposed two-level MLC and the similar system [29]. We assume 256QAM and the information rate is set to 6.0 bits per symbol. The inner polar code rates are chosen as 0.53 and 0.73 for the proposed scheme, and 0.68 and 0.82 for the scheme [29], for outer BCH and staircase codes, respectively. We assume the same polar length of 1024 bits for both schemes. Resulting symbol lengths are 1024 for the proposed system (regardless of outer codes), and 696 and 839 for the scheme [29] with outer BCH and staircase codes, respectively. From Fig. 13, it is observed that when the outer BCH code is employed, the error floors of both schemes are above the corresponding threshold. Also, the performance advantage of the proposed scheme becomes larger when the outer code rate is lower (the inner code rate is higher), since the ratio of the imperfect shaping ("Labeling1") becomes higher for the system in [29]. More specifically, when the outer staircase code is used, the proposed system outperforms the similar MLC scheme [29] by approximately 0.4 dB at the corresponding threshold. We note that the performance gap between the proposed system and the similar scheme [29] would become smaller as information rate increases since the optimal symbol distribution get closer to uniform, for which the impact of the imperfect shaping in [29] becomes smaller. Regarding complexity, the numbers of SD encodings and decodings required for the proposed scheme are about 68% and 82% of those for the scheme [29], for the outer BCH and staircase codes, respectively. Furthermore, the proposed system would be rather simpler in the sense that only a single bit-labeling scheme is used, whereas two different bit-labeling schemes are combined in [27]–[29].

VII. Conclusion

In this paper, we have proposed a novel probabilistic amplitude shaping (PAS) approach for concatenated two-level multi-level coding (MLC). By combining with PAS, the proposed concatenated two-level MLC scheme enables highly flexible transmission rate design by adjusting either or both of distribution matcher (DM) and FEC code rates. The proposed PAS

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/TCOMM.2022.3162365, IEEE Transactions on Communications

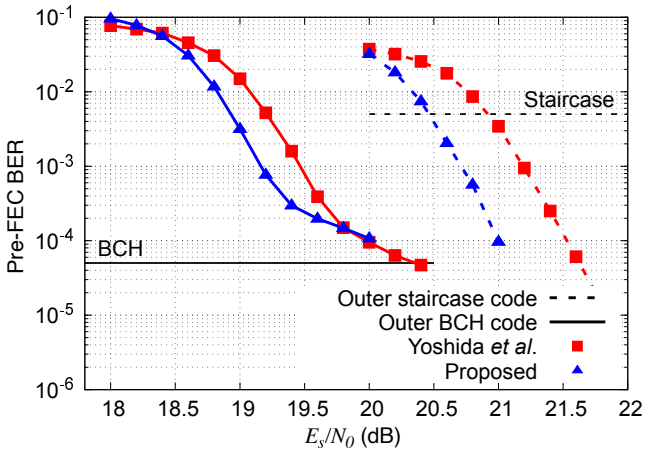


Fig. 13: Performance comparison with the similar MLC scheme [29] employing 256QAM at 6.0 bits per symbol.

is based on a new bit-labeling which is designed such that the coding gain of two-level MLC is maximized while achieving the target Maxwell-Boltzmann (MB) distribution. We used an approximated achievable information rate (AIR) of the proposed two-level MLC system with PAS based on which we optimized the balance of shaping and coding redundancies. We have demonstrated that the proposed MLC combined with PAS achieves a relative complexity reduction, e.g., 75% reduction in terms of the required number of SD decodings for 256QAM, while achieving even better performance than the conventional PAS in the simulations for a high information rate. Unlike existing PAS approaches for MLC, the proposed PAS scheme is capable of taking advantage of an excellent coding gain of two-level MLC without sacrificing shaping performance compared to the conventional PAS scheme.

In this paper, we have considered a two-level MLC scheme where only the LRB is encoded by a soft-decision (SD) FEC code. It is also possible to protect more than one bit level by an SD FEC code to reduce the error floor of the two-level MLC. However, in this case, the proposed PAS approach may not be directly applicable, and thus a new PAS scheme needs to be devised.

Appendix
Derivation of the Union Bound on the BER of the MRBs

As shown in Fig. 8, the union bound is a useful technique for estimating an error rate in the high SNR region. In what follows, we derive the union bound on the BER of the inner uncoded MRBs, that dominates the error floor of the proposed system. In the analysis, it is assumed that the inner decoder successfully decodes the LRB in the high SNR region, i.e., the BER of the LRB is zero.

Let $\mathcal{X}(B_{m-1})$ denote a subset of $\mathcal{X}$ whose $(m-1)$th bit label is $B_{m-1} \in \{0,1\}$ whose cardinality is $|\mathcal{X}(B_{m-1})| = 2^{m-1}$, and $x, \widehat{x} \in \mathcal{X}(b_{m-1})$ denote the transmitted and estimated symbols at the receiver side, respectively. Let $\mathcal{B}_{\text{MSB}}(x) \in \mathbb{F}_2^{m-1}$ denote the inverse of the symbol mapping function that maps a modulation symbol into a corresponding MRB label. Letting $d_{\text{H}}(\cdot,\cdot)$ denote the Hamming distance between two bit sequences, we define the coefficient that represents the ratio of incorrect bits in the estimated MRBs at the receiver side as

$$\alpha(x,\widehat{x}) \triangleq \frac{d_{\text{H}}(\mathcal{B}_{\text{MRB}}(x), \mathcal{B}_{\text{MRB}}(\widehat{x}))}{m-1}. \tag{16}$$

This coefficient depends on a bit-labeling scheme. Based on (16), the union bound on the BER of the inner uncoded MRBs is expressed as,

$$P_{\text{UB}}^{\text{MRBs}} = \sum_{B_{m-1}\in\mathbb{F}_2} P(B_{m-1}) \sum_{x\in\mathcal{X}(B_{m-1})} P(x|B_{m-1}) \times \sum_{\substack{\widehat{x}\in\mathcal{X}(B_{m-1}) \\ x\neq\widehat{x}}} \alpha(x,\widehat{x}) P(x\to\widehat{x}) \tag{17}$$

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

$$P(x\to\widehat{x}) = Q\left(\sqrt{\frac{|x-\widehat{x}|^2}{4\sigma^2}}\right), \tag{18}$$

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. Note that the LRB is uniformly distributed over $\mathbb{F}_2$, i.e., $P(B_{m-1}=0) = P(B_{m-1}=1) = 1/2$. When a transmit symbol $x \in \mathcal{X}(B_{m-1})$ has uniform distribution (without PAS), the conditional probability is $P(x|B_{m-1}) = 1/2^{m-1}$. However, $x \in \mathcal{X}(B_{m-1})$ is generally non-uniformly distributed for PAS systems. In the proposed MLC, the conditional probability of $x \in \mathcal{X}(B_{m-1})$ is independent of the LRB, i.e., $P(x|B_{m-1}=0) = P(x|B_{m-1}=1)$, since the LRB affects only the sign of PAM as shown in Fig. 3.

Acknowledgement

The authors would like to thank the editor and anonymous reviewers for their comments which have improved the paper.

References

[1] 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.

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] Z. Qu and I. B. Djordjevic, "On the probabilistic shaping and geometric shaping in optical communication systems," *IEEE Access*, vol. 7, pp. 21 454–21 464, 2019.

[17] 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.

[18] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 430–434, 2016.

[19] 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.

[20] 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.

[21] S. Chandrasekhar, B. Li, J. Cho, X. Chen, E. Burrows, G. Raybon, and P. Winzer, "High-spectral-efficiency transmission of PDM 256-QAM with parallel probabilistic shaping at record rate-reach trade-offs," in *Proc. 2016 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2016.

[22] A. Ghazisaeidi, I. F. de Jauregui Ruiz, R. Rios-Muller, L. Schmalen, P. Tran, P. Brindel, A. C. Meseguer, Q. Hu, F. Buchali, G. Charlet *et al.*, "65Tb/s transoceanic transmission using probabilistically-shaped PDM-64QAM," in *Proc. 2016 European Conference on Optical Communication (ECOC)*, pp. 1–3, 2016.

[23] J. Renner, T. Fehenberger, M. P. Yankov, F. Da Ros, S. Forchhammer, G. Böcherer, and N. Hanik, "Experimental comparison of probabilistic shaping methods for unrepeated fiber transmission," *J. Lightw. Technol.*, vol. 35, no. 22, pp. 4871–4879, 2017.

[24] J. Cho, X. Chen, S. Chandrasekhar, G. Raybon, R. Dar, L. Schmalen, E. Burrows, A. Adamiecki, S. Corteselli, Y. Pan *et al.*, "Trans-atlantic field trial using high spectral efficiency probabilistically shaped 64-QAM and single-carrier real-time 250-Gb/s 16-QAM," *J. Lightw. Technol.*, vol. 36, no. 1, pp. 103–113, 2018.

[25] Y. C. Gültekin, W. J. van Houtum, A. G. Koppelaar, and F. M. Willems, "Enumerative sphere shaping for wireless communications with short packets," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1098–1112, 2019.

[26] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, 2004.

[27] K. Sugitani, Y. Koganei, T. Oyama, and H. Nakashima, "Partial multilevel 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.

[28] 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.*, 2021.

[29] 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.

[30] 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.

[31] ETSI EN 302 307 V1.2.1, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications (DVB-S2)," Aug. 2009.

[32] P. Elias, "Error-free coding," *Trans. IRE Professional Group on Inf. Theory*, vol. 4, no. 4, pp. 29–37, 1954.

[33] 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.

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

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

[36] T. Mehmood, M. P. Yankov, A. Fisker, K. Gormsen, and S. Forchhammer, "Rate-adaptive concatenated polar-staircase codes for data center interconnects," in *Proc. Optical Fiber Communications Conference (OFC)*, pp. 1–3, 2020.

[37] 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, 2021.

[38] T. Fehenberger, D. S. Millar, T. Koike-Akino, K. Kojima, and K. Parsons, "Multiset-partition distribution matching," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 1885–1893, Nov. 2018.

[39] 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.

[40] 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.

[41] J. Cho, "Prefix-free code distribution matching for probabilistic constellation shaping," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 670–682, 2019.

[42] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[43] G. Böcherer, "Achievable rates for probabilistic shaping," *arXiv preprint arXiv:1707.01134v5*, 2018.

[44] A. Sheikh, A. Graell i Amat, G. Liva, and F. Steiner, "Probabilistic amplitude shaping with hard decision decoding and staircase codes," *J. Lightw. Technol.*, vol. 36, no. 9, pp. 1689–1697, 2017.

[45] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[46] A. Guillén i Fàbregas, A. Martinez, and G. Caire, *Bit-interleaved coded modulation*. Foundations and Trends in Communications and Information Theory, 2008, vol. 5, no. 1-2.

[47] A. Ingber and M. Feder, "Capacity and error exponent analysis of multilevel coding with multistage decoding," in *Proc. 2009 IEEE International Symposium on Information Theory (ISIT)*, pp. 1799–1803, 2009.

[48] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. M. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, 2009.

[49] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *Proc. 2010 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2010.

[50] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[51] ——, "Systematic polar coding," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 860–862, 2011.

[52] G. Sarkis, I. Tal, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Flexible and low-complexity encoding and decoding of systematic polar codes," *IEEE Trans. Commun.*, vol. 64, no. 7, pp. 2732–2745, 2016.

[53] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[54] ——, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, 2013.

[55] A. Elkelesh, M. Ebada, S. Cammerer, and S. ten Brink, "Decoder-tailored polar code design using the genetic algorithm," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 4521–4534, 2019.
[56] L. Huang, H. Zhang, R. Li, Y. Ge, and J. Wang, "AI coding: Learning to construct error correction codes," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 26–39, 2019.

Toshiki Matsumine (Member, IEEE) received the M.E. and Ph.D. degrees in information and communication engineering from Yokohama National University, Yokohama, Japan, in 2017 and 2020, respectively. In 2019, he held a research internship at Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. From 2020 to 2021, He was a postdoctoral researcher at the Technical University of Denmark, Lyngby, Denmark. He is currently a specially appointed assistant professor at Yokohama National University, Yokohama, Japan.

Metodi Plamenov Yankov (Member, IEEE) received the B. Eng. degree in the field of radio communications from the Technical University of Sofia, Bulgaria, in 2010, the M. Sc. degree in the area of signals and transmission technology for telecommunications from the Technical University of Denmark (DTU), Lyngby, in 2012, and the Ph.D. degree in the topic of constellation constrained capacity estimation near-capacity achieving digital methods from the Technical University of Denmark, Lyngby, Denmark. He held a Postdoc position with the DNRF Centre of Excellence SPOC from 2015 to 2017 and an Industrial Postdoc position with Fingerprint Cards A/S, Denmark, from 2017 to 2019. Since, he has been employed with DTU Fotonik, the Coding and Visual Communications group as a researcher. His research interests include among others information and communication theory of digital transceivers, digital signal processing techniques in general and for wireless and optical fiber communications in particular, machine learning, forward error correction codes, and information theory of biometrics. He is currently a member of the IEEE Photonics society.

Tayyab Mehmood (Student Member, IEEE) received the M.Sc. degree from the National University of Science & Technology, Islamabad, Pakistan in 2016, in the area of optical signal processing for radio over fiber systems and a Ph.D. degree from the Technical University of Denmark (DTU), Lyngby, Denmark in 2022 on the topic of rate-adaptive concatenated forward error correction for high throughput communications. Currently, he has been employed at Zeuxion ApS, Værløse, Denmark as an ASIC/FPGA engineer. His research interests include channel coding, information theory, and FPGA design for optical transport networks.

Søren Forchhammer (Member, IEEE) received the M.S. degree in engineering and the Ph.D. degree from the Technical University of Denmark, Lyngby, in 1984 and 1988, respectively. Currently, he is a Professor with DTU Fotonik, Technical University of Denmark, where he has been since 1988. He is Head of the Coding and Visual Communication Technology Group at DTU Fotonik and Flagship lead on Coding and Information theory of the DNRF CoE SPOC on optical communication. His main interests include, signal and image processing for communication, source coding, information theory, signal processing and coding for optical communication, distributed coding, visual communication technology