

# Multilevel Coding with General Decoding Metrics and Rateless Transmission

Trung Thanh Nguyen, *Student Member, IEEE*, and Lutz Lampe, *Senior Member, IEEE*

**Abstract**—Multilevel coding (MLC) is the main contender to the celebrated bit-interleaved coded modulation (BICM) technique for combining binary error-control codes with multilevel constellations. Although MLC has a more complex encoding-decoding structure, it can achieve a larger rate in a number of important scenarios such as multiple-input multiple-output (MIMO) and orthogonal modulation transmission. In this paper, we consider two issues related to the application of MLC. First, we examine the use of general decoding metrics in MLC, including mismatched metrics that arise from approximations to reduce detection complexity. We make use of recent advances in the analysis of BICM and apply those techniques to individual MLC transmission layers. Our contributions include rate analysis and metric-mismatch correction to improve throughput performance of MLC. Second, we consider the combination of MLC with binary rateless codes. Such a combination eliminates the need to carefully design code rate for each MLC layer. In slow fading environments, rateless coding can also seamlessly adapt to the instantaneous channel quality and achieve an increased average throughput compared to a fixed-rate MLC transmission. However, due to the MLC structure, we show that a naive combination of MLC and rateless coding can cause a significant rate loss. Thus, we propose a novel rotation rateless scheme which preserves the rate advantage of MLC over BICM. We provide relevant examples with MIMO, frequency-shift keying (FSK), and pulse-position modulation (PPM) signaling to demonstrate that our scheme can achieve throughput gains compared to BICM in a variety of transmission scenarios.

**Index Terms**—Multilevel coding (MLC), reduced-layer MLC, bit-interleaved coded modulation (BICM), mismatched decoding, generalized mutual information (GMI), rateless transmission.

## I. INTRODUCTION

The two primary resources of a communication system are channel bandwidth and transmit power [1]. Channel bandwidth utilization can often be increased with larger multilevel constellations. Transmit power requirement can be minimized by powerful error-control coding (ECC). However, non-binary codes for multilevel constellations are usually harder to design and implement, and hence binary codes are much more popular in coded transmission systems. A pragmatic approach to achieve both bandwidth and power efficiency is to combine binary ECC with multilevel constellation signaling. Multilevel coding (MLC) offers such a combination without

any reduction in the achievable rates compared to using non-binary codes [2], [3]. The main competing technique to MLC is bit-interleaved coded modulation (BICM) [4]–[8]. BICM also combines binary ECC with multilevel constellations, but in a simpler layout with just one instead of several coding layers and decoding stages as in MLC. The disadvantage of BICM over MLC is some rate loss. This loss can be small with Gray mappings [9] and thus, BICM has been widely employed in practice. Nevertheless, there are important cases where Gray mapping does not exist, or loses its relevance when the Euclidean-distance neighborhood of signal points is changed by the channel. For these cases, MLC might considerably outperform BICM. Examples of such cases are multiple-input multiple-output (MIMO) signaling for wireless communications [10], [11] and orthogonal modulation, which includes frequency-shift keying (FSK) for low-complexity detector implementation [12] and pulse-position modulation (PPM) for free-space optical (FSO) communications [13]. Considering that conventional MLC requires as many coding layers (and thus encoder-decoder pairs) as the number of levels, the concept of reduced-layer MLC (RL-MLC) [14], [15] has been introduced and demonstrated to provide good trade-off between achievable rates and the number of layers<sup>1</sup>. In particular, conventional MLC and BICM are special cases of RL-MLC where the number of layers equals the number of levels and one, respectively.

In this paper we make two distinct contributions to the family of MLC-based transmission schemes<sup>2</sup>. Our first contribution is a rate analysis and metric-mismatch correction for MLC using decoding metrics which are different from the maximum-likelihood (ML) metric. Such metrics are often used in practice to reduce detection complexity and/or in the case of channel uncertainty. We show that, for any generic decoding metric, the maximum achievable rate of MLC equals the sum of the maximum achievable rates of the layers. Furthermore, these layer rates can be estimated separately with the assumption that estimated data from lower layers is always correct. We then view each layer as a BICM scheme and adopt recent advances in the study of BICM with mismatched decoding metrics [6]–[8], [16] to the MLC case. In particular, we use the generalized mutual information (GMI) [17], [18] as an

Manuscript received November 13, 2010; revised February 23, 2011. This work was supported by the National Sciences and Engineering Research Council (NSERC) of Canada. The material in this paper was presented in part at the 2009 IEEE Global Communications Conference (GLOBECOM), Honolulu, Hawaii, USA.

The authors are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada. (e-mail: trungh@ece.ubc.ca, Lampe@ece.ubc.ca).

<sup>1</sup>We would like to stress the difference between the terms “level” and “layer.” Each level corresponds to a bit position in the labels of the signal points. Each layer corresponds to an encoder-decoder pair, and may be comprised of one or multiple levels.

<sup>2</sup>Without further clarification, by “MLC” we refer to the general RL-MLC scheme. When the reduced-layer nature is material in a discussion, we will make it explicit.

indication of the maximum achievable rate, and apply metric manipulations from [8], [16] to improve the performance of MLC.

Recently, the notion of rateless transmission has gained much attention in the communications community, e.g., [19], [20]. In the context of MLC, rateless transmission offers at least two attractive features. Firstly, it would eliminate the need for carefully tailoring code rate for each layer, e.g. according to the capacity design rule [3, Sec. IV-A]. Secondly, for transmission over time-varying channels, rateless codes seamlessly adapt to the instantaneous channel quality and attain a larger average throughput compared to channel adaptation by switching between fixed-rate codes. In our second contribution, we propose a novel rateless MLC scheme, which is compatible with any decoding metric and requires only one binary rateless code. We illustrate that combining rateless codes with MLC is not as straightforward as it is with BICM; and that a careless combination might lead to a considerable rate loss compared to fixed-rate MLC. We distinguish between coding loss due to the imperfectness of the code, and structural loss due to the structure of MLC. In our proposed scheme, the structural loss is non-zero only in so-called unstable systems. The stable or unstable nature of a rateless MLC scheme depends on the relationship between achievable rates of the layers. We illustrate that structural loss can be largely alleviated by controlling the acknowledgment delay from the receiver, and in practice a simple minimum segment-length (MSL) control rule can be effective enough in reducing this loss and maintaining the rate advantage of MLC.

The remainder of this paper is organized as follows. In Section II, we briefly describe MLC transmission. Our approach to rate analysis and metric-mismatch correction is presented in Section III. In Section IV, we introduce and analyze the operation and control in the proposed rateless MLC scheme. Practical applications to MIMO and orthogonal modulation are presented in Section V. Numerical results demonstrate that our rateless MLC scheme achieves throughput gains compared to BICM in a variety of transmission scenarios. Section VI completes the paper with concluding remarks.

## II. MULTILEVEL CODING PRELIMINARIES

Consider a discrete-time memoryless channel with transmit symbol  $X$  from the discrete constellation  $\mathcal{X}$ , received symbol  $Y$  from the set  $\mathcal{Y}$ , and channel transition probability  $p_{Y|X}(y|x)$ . The probability mass function of the transmit symbol is  $p_X(x)$ . Let  $M$  be the cardinality of  $\mathcal{X}$  and  $m = \log M$  be the number of labeling bits, i.e., the number of levels (notation  $\log(\cdot)$  denotes base-2 logarithm). An  $m$ -level,  $\kappa$ -layer MLC configuration,  $\kappa \leq m$ , is best described by a vector  $h = [h_0 \dots h_{m-1}]$  with  $h_i \in \{0, \dots, \kappa - 1\}$  and  $h_j \leq h_i$  if  $j < i$  [15]. The element  $h_i = k$  indicates that level  $i$  is in layer  $k$ . Each layer corresponds to an encoder-decoder pair. The MLC transmitter with  $\kappa$  independent binary encoders  $\text{enc}_0, \dots, \text{enc}_{\kappa-1}$  is presented in Figure 1(a). The corresponding MLC receiver with  $\kappa$  binary decoders  $\text{dec}_0, \dots, \text{dec}_{\kappa-1}$  is presented in Figure 1(b). The variables  $v_k, \hat{v}_k$ , and  $\hat{u}_k$ ,  $k = 0, \dots, \kappa - 1$ , in Figure 1 denote realizations

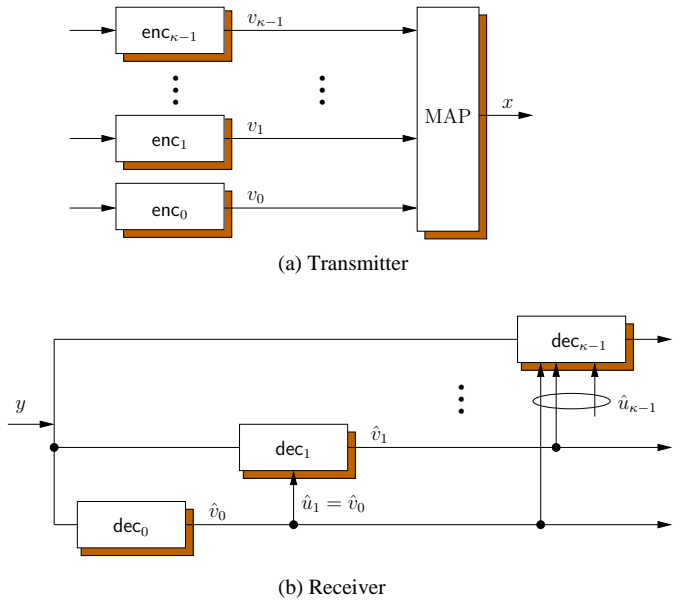


Fig. 1. MLC transmitter and receiver.

and decoding decisions of random variables explained in the following. Let  $B_0, \dots, B_{m-1}$  be the random variables of the labeling bits. Let  $V_k$  be the multivariate random variable of the labeling bits at layer  $k$ . That is, the elements of  $V_k$  are  $\{B_i : h_i = k\}$ . In  $V_k$  and other multivariate random variables below, we assume that the binary random variables  $\{B_i\}$  are ordered by their index  $i$ . At the receiver, the detection of  $B_i$  depends on the received symbol  $Y$  and the estimated value of bits from lower layers. Let  $D_i$  denote the multivariate random variable of these lower-layer bits, i.e., the elements of  $D_i$  are  $\{B_j : h_j < h_i\}$ . Furthermore, let  $b_i(X)$  and  $d_i(X)$  denote the value of  $B_i$  and  $D_i$  corresponding to the transmit symbol  $X$ . In conventional MLC [3] where each level is put in a separate layer, we have  $h = [0 \dots m - 1]$  and  $D_i = [B_0 \dots B_{i-1}]$ . In BICM where all levels are grouped into one layer, we have  $h = 0_{1 \times m}$  and  $D_i = \emptyset$ . Let  $U_k$  be the multivariate random variable of all lower-layer bits common to all levels at layer  $k$ . The elements of  $U_k$  are  $\{B_j : h_j < k\}$ .

As an example, consider a system with  $m = 3$  levels and  $\kappa = 2$  layers such that  $B_0$  is put in layer 0, and  $B_1$  and  $B_2$  are put in layer 1. Then,  $h = [0 \ 1 \ 1]$ ,  $V_0 = [B_0]$ ,  $V_1 = [B_1 \ B_2]$ ,  $U_0 = D_0 = \emptyset$ , and  $U_1 = D_1 = D_2 = [B_0]$ . That is, the detection of  $B_0$  depends only on  $Y$ , whereas the detection of  $B_1$  and  $B_2$  depends on  $Y$  and  $\hat{B}_0$ , which is the estimated value of  $B_0$ .

## III. MLC WITH GENERAL DECODING METRICS

In this section, we first establish that the achievable rate of MLC can be expressed in terms of the achievable rates of the layers (Section III-A). We then focus on MLC layers and provide expressions for the layer GMI (Section III-B). Based on these expressions, we propose a metric-mismatch correction technique for MLC (Section III-C). The latter two contributions are an extension of results from [6]–[8], [16] for BICM to MLC.

### A. Achievable Rates

Given a detection rule, we say rate  $C$  is achievable if for any  $\epsilon > 0$  and sufficiently large  $N$ , there exists a code  $\mathcal{C}(N', R)$  of length  $N' \geq N$  and rate  $R \geq C$  and a decoding algorithm  $\mathcal{D}$  such that the block error probability is less than or equal to  $\epsilon$ , cf. [21, Ch. 5], [22, Ch. 10]. We say  $C$  is the maximum achievable rate if it is the tightest upper bound on achievable rates.

*Theorem 1:* Given a detection rule, let  $C_k$  be an achievable rate of the  $k$ -th layer given decoding decisions from lower layers are correct,  $k = 0, \dots, \kappa-1$ . The MLC scheme achieves rate  $C = \sum_{k=0}^{\kappa-1} C_k$ . Furthermore, if each  $C_k$  is the *maximum* achievable rate of layer  $k$ , then  $C$  is the *maximum* achievable rate of the MLC scheme.

A proof of Theorem 1 is presented in the Appendix. We have the following remarks. First, as the theorem makes no assumption about detection rules being matched to the channel transition probabilities, it can be applied to mismatched decoding in the layers. Second, Theorem 1 allows us to consider each layer separately without concerning about error propagation when estimating achievable rates and manipulating mismatched metrics to increase these rates.

### B. Per-Layer Transmission and Generalized Mutual Information (GMI)

Since the maximum achievable rate of a mismatched decoding scheme is not known in general, we follow recent literature on BICM [6], [7] and use the GMI [17], [18] as an approximation to the layer maximum achievable rate.

The detector calculates bit metrics of the general form  $q_{B_i, Y, \hat{D}_i}(b, y, d)$  for  $B_i$  based on the received sample  $y \in \mathcal{Y}$  and estimated data from lower layers  $d \in \mathcal{B}^{|\mathcal{D}_i|}$ ,  $\mathcal{B} \triangleq \{0, 1\}$ . We assume  $q_{B_i, Y, \hat{D}_i}(b, y, d) > 0$ ,  $\forall b \in \mathcal{B}$ ,  $y \in \mathcal{Y}$ , and  $d \in \mathcal{B}^{|\mathcal{D}_i|}$ . It is also common for binary decoders to operate on the log-likelihood ratio (LLR) instead of bit metric values. The LLR is defined as

$$\Lambda_{q_{B_i, Y, \hat{D}_i}}(y, d) \triangleq \ln \frac{q_{B_i, Y, \hat{D}_i}(0, y, d)}{q_{B_i, Y, \hat{D}_i}(1, y, d)}. \quad (1)$$

For all levels  $i$  at layer  $k$ , we have  $\hat{D}_i = \hat{U}_k$ . Thus, the bit metric  $q_{B_i, Y, \hat{D}_i}(b, y, d)$  can also be written as  $q_{B_i, Y, \hat{U}_k}(b, y, u)$ . The layer metric is defined as

$$q_{V_k, Y, \hat{U}_k}(v, y, u) = \prod_{B_i \in V_k} q_{B_i, Y, \hat{U}_k}(b, y, u), \quad (2)$$

with  $v \in \mathcal{B}^{|V_k|}$  and  $u \in \mathcal{B}^{|U_k|}$ .

From Theorem 1, in the following calculations we assume that the estimated data from lower layers is always the same as transmitted. That is, we use  $D_i$  instead of  $\hat{D}_i$  and  $U_k$  instead of  $\hat{U}_k$ . For the  $i$ -th level and bit metric  $q_{B_i, Y, D_i}(b, y, d)$ , the

binary I-curve is defined as [6], [8], [17], [21]

$$\begin{aligned} I_{q_{B_i, Y, D_i}}(s) & \triangleq -\mathcal{E}_{B_i, Y, D_i} \left\{ \log \sum_{b \in \mathcal{B}} p_{B_i}(b) \left[ \frac{q_{B_i, Y, D_i}(b, Y, D_i)}{q_{B_i, Y, D_i}(B_i, Y, D_i)} \right]^s \right\} \quad (3) \\ & = -\mathcal{E}_{X, Y} \left\{ \log \sum_{b \in \mathcal{B}} p_{B_i}(b) \left[ \frac{q_{B_i, Y, D_i}(b, Y, d_i(X))}{q_{B_i, Y, D_i}(b_i(X), Y, d_i(X))} \right]^s \right\}. \quad (4) \end{aligned}$$

The binary GMI of the level is the peak value of this curve,

$$I_{q_{B_i, Y, D_i}}^{\text{gmi}} \triangleq \max_{s > 0} I_{q_{B_i, Y, D_i}}(s). \quad (5)$$

For matched metrics, i.e., when  $q_{B_i, Y, D_i}(b, y, d)$  is proportional to the transition probability  $p_{Y|B_i, D_i}(y|b, d)$ , the binary I-curve peaks at  $s = 1$  and the GMI equals the average mutual information  $I(B_i; Y|D_i)$ .

With uniform input, from (1) and (4), the binary I-curve can be expressed in terms of the LLR as

$$\begin{aligned} I_{q_{B_i, Y, D_i}}(s) & = 1 - \\ & \mathcal{E}_{X, Y} \left\{ \log(1 + \exp(-\text{sgn}(b_i(X)) \Lambda_{q_{B_i, Y, D_i}}(Y, d_i(X))s)) \right\}, \quad (6) \end{aligned}$$

where the function  $\text{sgn}(\cdot)$  is defined for the labeling bits as  $\text{sgn}(0) = 1$  and  $\text{sgn}(1) = -1$ .

Corresponding to the layer metric  $q_{V_k, Y, U_k}(v, y, u)$ , the layer I-curve is defined as

$$\begin{aligned} I_{q_{V_k, Y, U_k}}(s) & \triangleq \\ & -\mathcal{E}_{V_k, Y, U_k} \left\{ \log \sum_{v \in \mathcal{B}^{|V_k|}} p_{V_k}(v) \left[ \frac{q_{V_k, Y, U_k}(v, Y, U_k)}{q_{V_k, Y, U_k}(V_k, Y, U_k)} \right]^s \right\}. \quad (7) \end{aligned}$$

Similarly to the transition from (3) to (4), we have

$$\begin{aligned} I_{q_{V_k, Y, U_k}}(s) & = -\mathcal{E}_{X, Y} \\ & \left\{ \log \sum_{v \in \mathcal{B}^{|V_k|}} p_{V_k}(v) \left[ \frac{q_{V_k, Y, U_k}(v, Y, u_k(X))}{q_{V_k, Y, U_k}(v_k(X), Y, u_k(X))} \right]^s \right\}. \quad (8) \end{aligned}$$

From (4), (2), and (8), it can be shown that the layer I-curve equals the sum of the binary I-curves of the levels,

$$I_{q_{V_k, Y, U_k}}(s) = \sum_{B_i \in V_k} I_{q_{B_i, Y, D_i}}(s). \quad (9)$$

The layer GMI is the peak value of this layer I-curve,

$$I_{q_{V_k, Y, U_k}}^{\text{gmi}} \triangleq \max_{s > 0} I_{q_{V_k, Y, U_k}}(s). \quad (10)$$

By Theorem 1, the MLC scheme achieves a rate equal to the sum of the layer GMIs.

### C. Metric-Mismatch Correction for MLC

The layer GMI (10) is less than or equal to the sum of the binary GMIs (5) in that layer. Equality is achieved if and only if the binary I-curves are *harmonic*, that is, when they attain their peaks at the same  $s$ -coordinate [8]. Let  $s_{q_{B_i, Y, D_i}}$  be the *critical point* (the  $s$ -coordinate of the peak) of  $I_{q_{B_i, Y, D_i}}(s)$ . Then, a new metric

$$q'_{B_i, Y, D_i}(b, y, d) = [q_{B_i, Y, D_i}(b, y, d)]^{c_i} \quad (11)$$

with  $c_i = s_{q_{B_i, Y, D_i}}/s^*$  for any  $s^* > 0$  yields a binary I-curve with the same GMI, but with the new critical point  $s_{q'_{B_i, Y, D_i}} = s^*$  [8]. We propose applying this metric correction to make the binary I-curves harmonic and thus improve the layer GMI. It has also been demonstrated in [8] that while the choice of  $s^*$  does not affect the GMI and the error-rate performance of word and max-product (a.k.a. min-sum) symbol-by-symbol (SBS) decoding, the specific value  $s^* = 1$  is preferable for sum-product SBS decoding.

Since raising the bit metric to a power corresponds to scaling the LLR with the same factor, adoption of the above metric-mismatch correction for MLC results in the LLR scaling

$$\Lambda_{q'_{B_i, Y, \hat{D}_i}}(y, d) = s_{q_{B_i, Y, D_i}} \Lambda_{q_{B_i, Y, \hat{D}_i}}(y, d). \quad (12)$$

We would like to stress that the scaling factor  $s_{q_{B_i, Y, D_i}}$  in (12) is obtained without consideration to possible errors at lower layers.

Finally, we note that the layer GMIs might further be increased by post-detection processing of LLRs using the scalar correction from [16] or vector-correction functions from [8], which however come with increased processing complexity.

## IV. RATELESS MULTILEVEL CODING

We start this section with the description of our novel rateless MLC scheme. Then, we provide an operation analysis and devise a control scheme to reduce the structural rate loss when necessary. Our scheme is compatible with any mismatched decoding in the layers. Following Theorem 1 and for generality, we continue to use the notations  $C_k$ ,  $k = 0, \dots, \kappa - 1$ , and  $C = \sum_{k=0}^{\kappa-1} C_k$  to denote the maximum achievable rates of the layers (assuming no error from lower layers), and of MLC, respectively. We will resort to the GMI as an approximation to the maximum achievable rates again when showing numerical results in Section V.

### A. Encoding and Decoding

In rateless BICM, the binary encoder takes in a fixed-length block of message bits and continuously produces coded bits, which are mapped to transmit symbols. The receiver can produce metric samples immediately after each received symbol. These metric samples are accumulated until they can provide enough information for successful decoding. At this point, the receiver acknowledges the transmitter and the transmitter switches to a new message block, cf. [20]. We now consider a naive combination of rateless coding and MLC, in which the encoder-decoder pair at each layer (see Figure 1) is implemented in a rateless-coding fashion. At layer

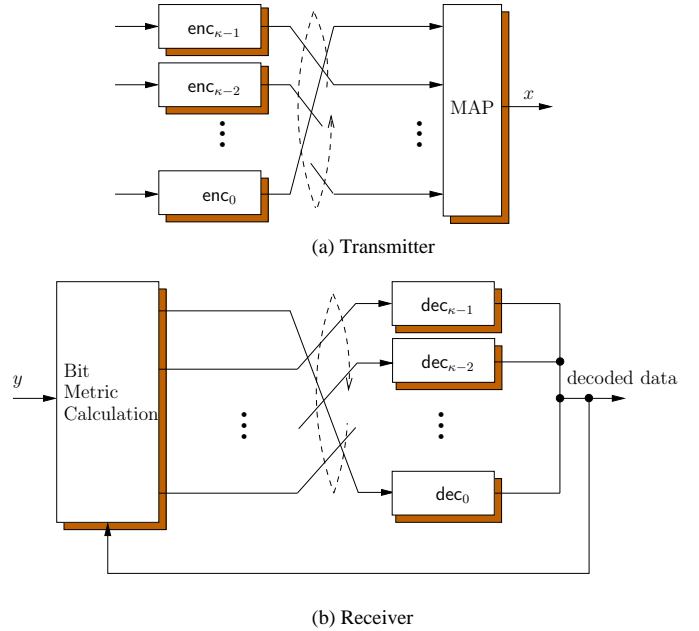


Fig. 2. Rotation rateless MLC transmitter and receiver. Stream redirection occurs at switching to a new interval.

0, metric samples become available to the decoder right after each received symbol, as in rateless BICM. However, at any layer  $k > 0$ , the decoder collects symbol samples and waits for estimated data from lower layers. Only then, metric samples can be calculated. The estimated data from lower layers does not become available right after each received symbol, but in large blocks. Thus, by the time this data becomes available, the segment of metric samples at layer  $k$  may have already accumulated more information than necessary for successful decoding. This means that layer  $k$  is underutilized and the exceeding channel use is wasted. It is only at the lowest layer that we are able to collect just enough information for successful decoding and avoid this underutilization.

The above observation leads us to propose the following scheme. The transmitter employs  $\kappa$  binary rateless encoders  $enc_k$ ,  $k = 0, \dots, \kappa - 1$ . In the naive combination above, output from encoder  $enc_k$  is always mapped to layer  $k$ . In our proposed scheme as illustrated in Figure 2(a), output from  $enc_k$  is mapped to different layers at different time intervals. More specifically,  $enc_k$  cyclically directs its output through layers  $(\kappa - 1), (\kappa - 2), \dots, 0$  during  $\kappa$  time intervals  $t - (\kappa - 1), t - (\kappa - 2), \dots, t$ . Each time interval corresponds to the transmission of a number of symbols. The number of symbols in each interval might vary from one interval to another, and is discussed in detail below.

The corresponding receiver is shown in Figure 2(b). Each binary decoder  $dec_k$ ,  $k = 0, \dots, \kappa - 1$ , collects information from all layers, starting from the highest layer  $\kappa - 1$  and ending at the lowest layer 0. The decoders attempt to decode only when collecting information from layer 0. Thus, during any time interval, only the decoder collecting information from layer 0 attempts to decode.

Figure 3(a) shows the information segments at the receiver over  $\kappa$  time intervals. Suppose some decoder  $dec_k$  is collecting

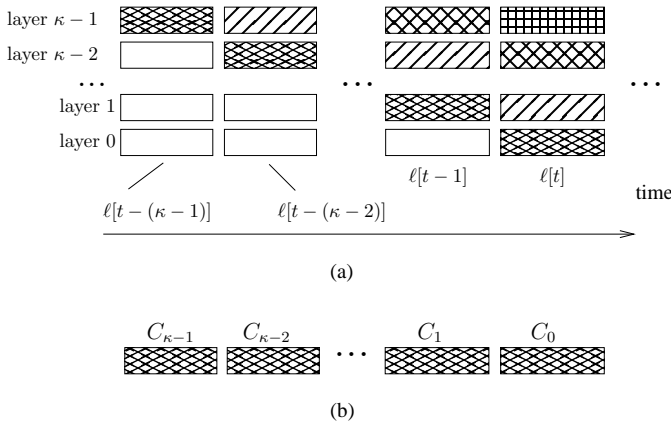


Fig. 3. (a) Information segments at the receiver. Segments with the same shading belong to the same codeword of a binary encoder-decoder pair. White segments are already decoded information. Segments may have different lengths. Among the  $\kappa$  binary decoders, during any time interval, only the one that is collecting information from layer 0 attempts to decode. (b) Each binary encoder-decoder pair experiences a time-varying channel with different interval maximum achievable rates  $C_i$ ,  $i = 0, \dots, \kappa - 1$ .

information from layer 0 during interval  $t$ . This decoder has also collected information segments from layers  $(\kappa - 1), \dots, 1$  in the preceding  $(\kappa - 1)$  intervals. When  $\text{dec}_k$  gathers that it might have accumulated enough information, it makes a decoding attempt. If decoding fails,  $\text{dec}_k$  continues to collect information from layer 0, and attempts decoding again. Eventually, decoding will be successful, which marks the end of this time interval. The receiver then sends an acknowledgment to the transmitter. Upon receiving the acknowledgment, the transmitter: (i) changes the input of the corresponding encoder  $\text{enc}_k$  to a new message block, and (ii) re-maps the encoded bit streams to different layers in the cyclical manner as illustrated in Figure 2(a). That is, the stream that was mapped to layer  $k$  is now mapped to layer  $(k - 1)$ , for  $k = 1, \dots, \kappa - 1$ , and the stream that was mapped to layer 0 is now mapped to layer  $\kappa - 1$ . In the next time interval  $t + 1$ , the decoders will collect information from different layers, corresponding to this new streaming. From the viewpoint of a fixed encoder-decoder pair  $(\text{enc}_k, \text{dec}_k)$ , the overall channel is segment-wise time varying. Each segment is associated with a different maximum achievable rate. This is illustrated in Figure 3(b).

The proposed scheme requires an initialization. This can be seen in Figure 3(a), where white segments represent already decoded information. Therefore, at the start of the transmission session, the transmitter sends some preamble data that is known to the receiver in the corresponding segments. This preamble transmission appears only once and thus the associated rate loss will be negligible for sufficiently long sessions.

If all binary encoder-decoder pairs use the same rateless code, then there is no need to have  $\kappa$  different decoders as shown in Figure 2(b), but a single binary decoder will be sufficient. Similarly, all layers can share one binary encoder hardware in a time-multiplexing fashion, at the cost of reduced encoding speed.

## B. Operation Analysis and Control

As shown in Figure 3, let  $\ell[t] \geq 0$  be the number of transmit symbols during time interval  $t$ . While  $\ell[t]$  is an integer number, for simplicity we let  $\ell[t]$  assume real values in the following analysis; the effect of this relaxation becomes negligible for long codewords. By the end of time interval  $t$ , one codeword has been successfully decoded. We distinguish two variables related to the transmission of this codeword:  $\theta[t]$  is the nominal amount of information collected by the receiver, i.e.,

$$\theta[t] = \sum_{k=0}^{\kappa-1} C_k \ell[t - k], \quad (13)$$

and  $\nu[t]$  is the minimal amount of information sufficient for successful decoding. Since we assume successful decoding, we have  $\nu[t] \leq \theta[t]$ .

We assume all message blocks input to the binary encoders have the same length of  $K$  bits. We call a binary rateless code *ideal* if  $\nu[t] = K$ . Practical codes, e.g. Raptor codes [20], require positive overheads and thus have a positive average  $(\nu[t] - K)$ . We call this value the *coding loss* as it measures the required overhead of the code. The difference between  $\theta[t]$  and  $\nu[t]$  is due to the MLC structure. We therefore call the average of  $(\theta[t] - \nu[t])$  the *structural loss*. The total rate loss of an MLC scheme, i.e. the gap between the average throughput and  $C$ , is the sum of its coding and structural loss.

If successful decode does not occur at the start of the interval when  $\ell[t] = 0$ , we can collect just enough additional information with  $\ell[t] > 0$  and terminate decoding such that  $\theta[t] = \nu[t]$ . That is, no structural loss incurs. However, if successful decoding is possible at  $\ell[t] = 0$ , we have likely accumulated more than enough information and wasted channel resources. Therefore, if  $\ell[t]$  varies around an *equilibrium length* and  $\ell[t] > 0$ , we assume that the transmission has no structural loss. In the following we will elaborate on this equilibrium length and the conditions under which the system is stable, that is, when  $\ell[t]$  automatically converges to and remains around the equilibrium length.

1) *Equilibrium Length*: In systems with no structural loss, we have  $\theta[t] = \nu[t]$ . From (13),

$$\ell[t] = -\frac{C_{\kappa-1}}{C_0} \ell[t - (\kappa - 1)] - \dots - \frac{C_1}{C_0} \ell[t - 1] + \frac{1}{C_0} \nu[t].$$

Let  $n[t] = [\ell[t - (\kappa - 1)] \dots \ell[t - 1]]^T$  be a vector of  $(\kappa - 1)$  non-negative previous transmission lengths that represents the receiver state at the beginning of interval  $t$ . For  $t \geq t_{\text{initial}} = 0$ , the state therefore evolves according to the linear time-invariant state-space equation

$$n[t + 1] = An[t] + b\nu[t], \quad (14)$$

with

$$A = \begin{bmatrix} 0_{(\kappa-2) \times 1} & & I_{\kappa-2} \\ -\frac{C_{\kappa-1}}{C_0} & -\frac{C_{\kappa-2}}{C_0} & \dots & -\frac{C_1}{C_0} \end{bmatrix}$$

and

$$b = \begin{bmatrix} 0_{(\kappa-2) \times 1} \\ \frac{1}{C_0} \end{bmatrix},$$

where  $I_{\kappa-2}$  is the identity matrix of size  $(\kappa-2) \times (\kappa-2)$  and  $0_{(\kappa-2) \times 1}$  is the all-zero column vector of length  $(\kappa-2)$ .

Let  $\bar{\nu}$  be the mean value of  $\nu[t]$ . We call the solution of the equation

$$n[t+1] = An[t] + b\bar{\nu}, \quad (15)$$

the equilibrium point. This equilibrium point is found as

$$n_e = \bar{\nu}(I_{\kappa-1} - A)^{-1}b = \ell_e 1_{(\kappa-1) \times 1}, \quad (16)$$

where  $\ell_e$  is the equilibrium length

$$\ell_e = \frac{\bar{\nu}}{\sum_{i=0}^{\kappa-1} C_i} = \frac{\bar{\nu}}{C}, \quad (17)$$

and  $I_{\kappa-1}$  is the identity matrix of size  $(\kappa-1) \times (\kappa-1)$ . We observe that:

- If  $\nu[t]$  is a constant, i.e.  $\nu[t] = \bar{\nu}$ , once at the equilibrium point, the system stays there.
- With ideal codes, i.e.  $\nu[t] = K$ , once at the equilibrium point, the system stays there and suffers zero total loss. That is, with ideal codes, we can achieve a throughput equal to  $C$ .

2) *Stable Systems*: The system is stable if the eigenvalues of the matrix  $A$  in (14) lie inside the unit circle in the complex plane. In this case, the state  $n[t]$  will automatically hover around the equilibrium point  $n_e$ . Then, the interval length  $\ell[t]$  stays close to  $\ell_e$  and is positive, so that  $\theta[t] = \nu[t]$  holds and the transmission has zero structural rate loss. For stable systems, the only loss with respect to  $C$  is the coding loss and our proposed MLC scheme is an optimal way of combining binary rateless codes and multilevel constellations.

The eigenvalues of  $A$  are the roots  $r_i$  of the polynomial

$$U(z) = 1 + \frac{C_1}{C_0}z^{-1} + \frac{C_2}{C_0}z^{-2} + \dots + \frac{C_{\kappa-1}}{C_0}z^{\kappa-1}. \quad (18)$$

There are several methods to determine whether  $|r_i| < 1$  for all roots of (18) [23, Ch. 4.5]. For the special cases of  $\kappa = 2$  and  $\kappa = 3$  layers, we can make the criteria for stability explicit:

- $\kappa = 2$ :  $U(z) = 1 + \frac{C_1}{C_0}z^{-1}$ , and the root  $r_1 = -C_1/C_0$  lies inside the unit circle if and only if  $C_0 > C_1$ . In other words, a two-layer scheme is stable if the lower layer has larger maximum achievable rate than the upper layer.
- $\kappa = 3$ :  $U(z) = 1 + \frac{C_1}{C_0}z^{-1} + \frac{C_2}{C_0}z^{-2}$ , and the two roots  $r_{1,2} = \frac{1}{2C_0}(-C_1 \pm \sqrt{C_1^2 - 4C_0C_2})$  lie inside the unit circle if and only if ( $C_0 > C_2$  and  $C_0 + C_2 > C_1$ ).

3) *Unstable Systems*: If the linear system (14) is unstable, the interval length  $\ell[t]$  would oscillate with increasing amplitude. Due to the constraint  $\ell[t] > 0$ , the length  $\ell[t]$  will eventually swing between zero and some extreme values. This behavior is illustrated in Figure 4 for the example of  $\kappa = 3$ ,  $C_0 = 1$ ,  $C_1 = 1.7$ ,  $C_2 = 1.3$ , and thus  $C = 4$  bit per channel use (bpcu),  $\nu[t] = K$  for all  $t$  (the binary code is ideal), and the initial state  $n[0] = (K/C)[1.0 \ 1.2]^T$ . At large values of  $t$ , the system settles into the following pattern: there is one long interval of length  $K$ , followed by two intervals of length zero. In the long interval, one codeword is decoded with all samples from layer 0. In the next interval, one codeword is decoded with all samples from layer 1 and no more samples from layer 0 are needed (hence, the interval length is zero).

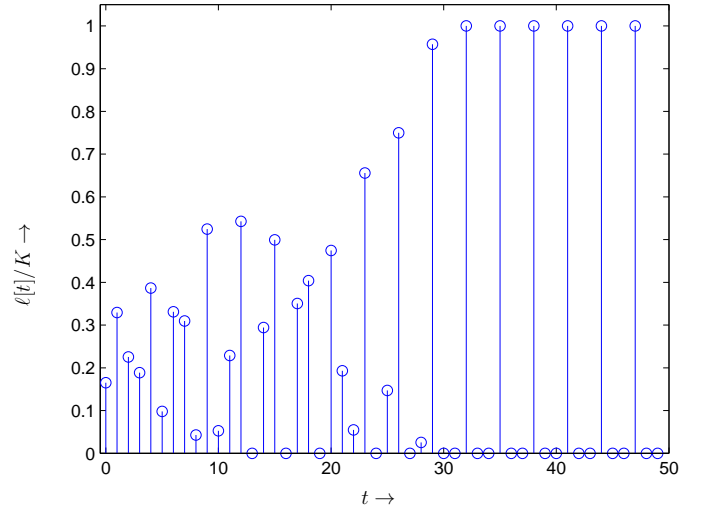


Fig. 4. Interval length  $\ell[t]$  of an unstable system.

Similarly, the next codeword is decoded with all samples from layer 2. Even though the rateless code is ideal, the system achieves a data rate of only 3 bpcu, which is well below  $C$ . This behavior calls for a control mechanism to minimize the structural loss for unstable systems.

4) *Control for Unstable Systems*: The receiver can delay sending the acknowledgment to the transmitter after successful decoding. By controlling this delay, we can reduce the structural loss. To illustrate the potential of such a control, we consider a *hypothetical scenario* where we can predict the values of  $\nu[t]$  in a near future. Suppose that at the beginning of time interval  $\tau$ , the receiver state is  $n[\tau]$ , and we predict the values of  $\nu[t]$  over the finite time horizon  $T$  to be  $\nu[\tau], \dots, \nu[\tau+T-1]$ . Beyond this horizon, we extrapolate that  $\nu[t]$  equals to the average of the predicted values  $\bar{\nu}[\tau]$ , given by

$$\bar{\nu}[\tau] = \frac{1}{T} \sum_{t=\tau}^{\tau+T-1} \nu[t]. \quad (19)$$

We observe from Eqns. (14) to (17) that, if  $\nu[t]$  remains constant as  $\nu[t] = \bar{\nu}[\tau]$  for  $t \geq \tau+T$ , the transmission suffers no structural loss if all the segment lengths equal  $\bar{\nu}[\tau]/C$ . Therefore, we want to settle into the state  $\ell[t] = \bar{\nu}[\tau]/C$  for  $t \geq \tau+T$ . We end up having an extended control period from  $\tau$  to  $\tau+T+\kappa-2$  with the desired ending state

$$\ell[\tau+T] = \dots = \ell[\tau+T+\kappa-2] = \frac{\bar{\nu}[\tau]}{C}. \quad (20)$$

For the period  $[\tau, \tau+T-1]$ , we plan the lengths  $\ell[\tau], \dots, \ell[\tau+T-1]$  to minimize the structural loss. The initial amount of information stored at the receiver at the beginning of time interval  $t = \tau$  (the initial information inventory) is

$$\mathcal{I}_{\text{beginning}} = \sum_{k=1}^{\kappa-1} \sum_{j=k}^{\kappa-1} C_j \ell[\tau-k]. \quad (21)$$

The ending information inventory at the end of time interval  $\tau+T+\kappa-2$  (or beginning of the time interval  $t = \tau+T+$

$\kappa - 1$ ), is

$$\mathcal{I}_{\text{end}} = \sum_{k=1}^{\kappa-1} \sum_{j=k}^{\kappa-1} C_j \frac{\bar{\nu}[\tau]}{C}. \quad (22)$$

During the control period, the total amount of arriving information is  $C \sum_{t=\tau}^{\tau+T+\kappa-2} \ell[t]$ , and we spend  $\sum_{t=\tau}^{\tau+T+\kappa-2} \nu[t]$  to successfully decode  $T + \kappa - 1$  codewords. The structural loss is therefore

$$\mathcal{I}_{\text{structural loss}} = C \sum_{t=\tau}^{\tau+T+\kappa-2} \ell[t] - \sum_{t=\tau}^{\tau+T+\kappa-2} \nu[t] + \mathcal{I}_{\text{beginning}} - \mathcal{I}_{\text{end}}. \quad (23)$$

Since the beginning and ending receiver state are fixed, the structural loss is minimized if  $\sum_{t=\tau}^{\tau+T-1} \ell[t]$  is minimized. Planning the segment lengths can then be stated as the following linear programming (LP) problem:

$$\begin{aligned} \{\hat{\ell}[\tau] \dots \hat{\ell}[\tau + T - 1]\} &= \operatorname{argmin} \sum_{t=\tau}^{\tau+T-1} \ell[t] \\ \text{subject to:} \\ \sum_{k=0}^{\kappa-1} C_k \ell[t-k] &\geq \nu[t], \quad \forall t \in [\tau, \tau + T + \kappa - 2], \\ \left( \ell[\tau - (\kappa - 1)] \dots \ell[\tau - 1] \right)^T &= n[\tau], \\ \ell[\tau + T] = \dots = \ell[\tau + T + \kappa - 1] &= \bar{\nu}[\tau]/C, \\ \ell[t] &\geq 0. \end{aligned} \quad (24)$$

As the transmission progresses, we apply the model predictive control (MPC) technique, e.g. [24], to adjust to the increment of  $t$ . The MPC technique is summarized as follows:

- 1: At the beginning of the time interval  $t = \tau$ , predict  $\nu[\tau], \dots, \nu[\tau + T - 1]$ .
- 2: Plan the segment lengths for the period  $[\tau, \tau + T - 1]$  by solving (24).
- 3: Receive at least  $\hat{\ell}[\tau]$  symbols in this time interval.
- 4: Repeat for  $t \leftarrow \tau + 1$ .

An alternative control rule to the MPC technique above is to always receive at least  $\ell_{\min}$  symbols in each interval. We call this the minimal segment length (MSL) control rule. This simple rule comes from observing the behavior of unstable systems without control, as described in Section IV-B3. Since the decoders have to wait for their turn, the one that collects most of the information from a low-rate layer needs a relatively long interval and forces other decoders to accumulate too much information. By setting a proper minimum value for all  $\ell[t]$ , MSL reduces the long wait by making the total amount of information to be distributed more equally over the  $\kappa$  intervals.

We now illustrate the effectiveness of the control techniques via a hypothetical example. Let  $\nu[t]$  be a random process such that  $\nu[t] = (1.05 + \epsilon[t])K$ . The component  $\epsilon[t]$  is Rayleigh-distributed with mean 0.04. Therefore,  $\bar{\nu}/K = 1.09$  and thus the coding loss is 9%. This model approximates the simulation result of a Raptor code from [25, Fig. 6]. We consider the same MLC example in Section IV-B3. That is,  $\kappa = 3$ ,  $C_0 = 1$ ,  $C_1 = 1.7$ , and  $C_2 = 1.3$  bpcu. Without control, the transmission suffers a structural loss of 36%. For MPC with perfect prediction, the structural loss versus  $T$  is plotted in the upper half of Figure 5. It is interesting to observe that even with short  $T = 1$ , MPC reduces the structural

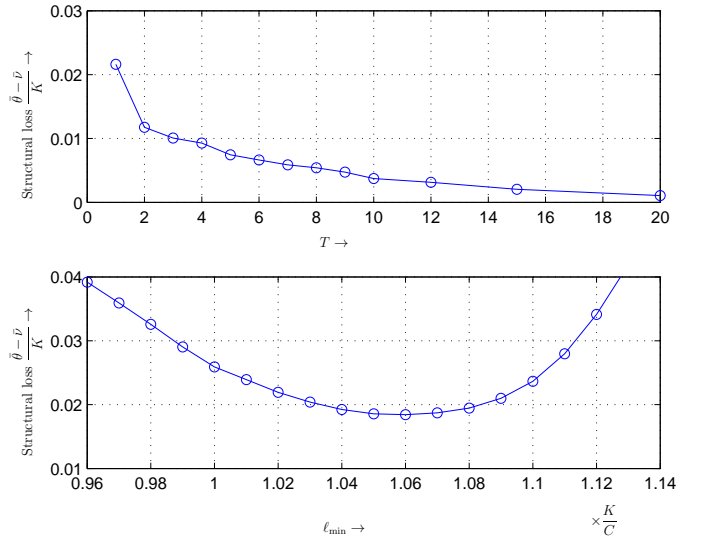


Fig. 5. Controlling structural loss in an unstable system with  $[C_0 \ C_1 \ C_2] = [1.0 \ 1.7 \ 1.3]$  bpcu. Top: structural loss vs.  $T$  for model predictive control (MPC) with perfect prediction. Bottom: structural loss vs.  $\ell_{\min}$  for minimum segment length (MSL) control.

loss to just about 2%. With long horizon  $T$ , it can almost completely alleviate the structural loss. Thus, controlling the acknowledgment delay can indeed reduce the structural loss. Unfortunately, perfect prediction of  $\nu[t]$  over a long horizon is rather idealistic. With the simple and practical MSL control rule, the structural loss versus  $\ell_{\min}$  is plotted in the lower half of Figure 5. The result shows that MLC can reduce the structural loss to a very small value of less than 2%. This value is much less than the coding loss. These results suggest that MSL control is indeed a practical way of stabilizing rateless MLC systems, and we apply it in the application examples below.

## V. APPLICATIONS

In this section, we demonstrate the efficacy of our rateless MLC scheme in a number of transmission scenarios. The examples include MIMO and orthogonal modulation transmission in stationary or slow fading channels, and using matched or approximate metrics. Following Section III-B, we use the GMI as an approximation to the maximum achievable rates of the layers. Furthermore, we refer to the sum of the layer GMIs as the MLC GMI. An off-the-shelf binary rateless Raptor code as described in [8, Example 1] with  $K = 9500$  is used in all simulations. We use sum-product SBS decoding by the belief-propagation algorithm in the log domain with a maximum of 200 iterations to decode the joint factor graph of the two Raptor component codes [26, Fig. 1b]. When a decoding attempt fails, an additional information amount of  $0.01K$  is collected before the next attempt.

### A. MIMO-QAM

Consider MIMO transmission with  $N_t$  transmit and  $N_r$  receive antennas. Each transmit antenna emits symbols from a constituent quadrature amplitude modulation (QAM) constellation  $\mathcal{A}$  of size  $M_A$ . Each symbol of the MIMO constellation

$\mathcal{X}$  is thus a vector  $x = [x_0 \dots x_{N_t-1}]^T$  of  $N_t$  elements  $x_i \in \mathcal{A}$ . The size of the multidimensional transmit constellation is  $M = (M_{\mathcal{A}})^{N_t}$ . Assuming a flat fading channel with the matrix of complex gains  $H$  and complex additive white Gaussian noise (AWGN) vector  $n$  with variance  $N_0$  per element, the received symbol is given by

$$y = Hx + n. \quad (25)$$

The channel transition probability follows as

$$p_{Y|X}(y|x) \propto \exp\left(-\frac{|y - Hx|^2}{N_0}\right). \quad (26)$$

It was shown in [10, Sec. IV-E] that there exists a significant gap between the BICM GMI and the constellation-constrained channel capacity  $I(X;Y)$ , even when Gray mappings are applied for the constituent constellation  $\mathcal{A}$ , and hence for  $\mathcal{X}$ . Thus, it is appealing to use MLC with MIMO signaling.

Let us consider the case of 4-QAM with  $N_t = 4$  transmit antennas and matched LLR metric

$$\Lambda_{q_{B_i,Y}, \hat{D}_i}(y, d) = \ln \frac{\sum_{x \in \mathcal{X}_{d_i,i}^{d,b}} p_{Y|X}(y|x)}{\sum_{x \in \mathcal{X}_{d_i,i}^{d,1}} p_{Y|X}(y|x)}, \quad (27)$$

where the notation  $\mathcal{X}_{d_i,i}^{d,b}$  denotes the set of  $x$  which have the labeling bits  $d_i(x) = d$  and  $b_i(x) = b$ ,  $b \in \mathcal{B}$ . The size of the constellation is  $M = 256$  and the number of labeling bits is  $m = 8$ . We use MLC with  $\kappa = 2$  layers. The first four bits are grouped into layer 0, and the other four bits are grouped into layer 1. We consider an i.i.d. Rayleigh fading channel and two scenarios of  $N_r = 2$  and  $N_r = 4$  receive antennas. For both scenarios, the GMI of layer 0 is less than that of layer 1, and therefore the system is unstable. We apply the MSL control with  $\ell_{\min} = 1.06K/C$ . The signal-to-noise ratio (SNR) is defined as  $E_r/N_0$ , where  $E_r$  is the average received power at each antenna, cf. [10]. The simulation results are plotted in Figure 6. It can be observed that MLC with just two layers can considerably reduce the gap between BICM GMI and the average mutual information  $I(X;Y)$ . Over the whole SNR range, the simulated throughput is about 92% of the MLC GMI. This ratio is close to the reported value in [25] for a Raptor code over a binary AWGN channel. Therefore, we conclude that most of the rate loss is the coding loss, and the structural loss is small. Overall, the results demonstrate a significant throughput gain from using our rateless MLC scheme instead of BICM.

## B. Orthogonal Modulation

In orthogonal modulation, the Euclidean distances between all signal points are identical. Hence, for  $M > 2$ , Gray mappings do not exist. We consider two prominent examples of orthogonal modulation, which are FSK and PPM.

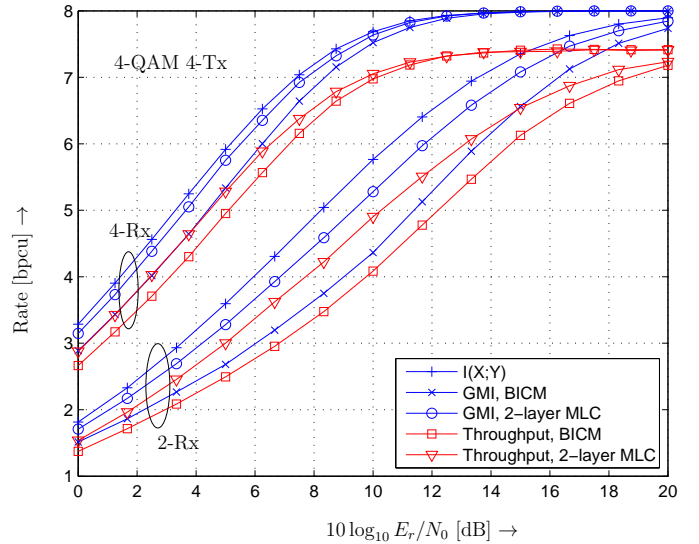


Fig. 6. GMI and throughput of rateless MLC for 4-QAM MIMO transmission with 4 transmit and 2 and 4 receive antennas over an i.i.d. Rayleigh fading channel.

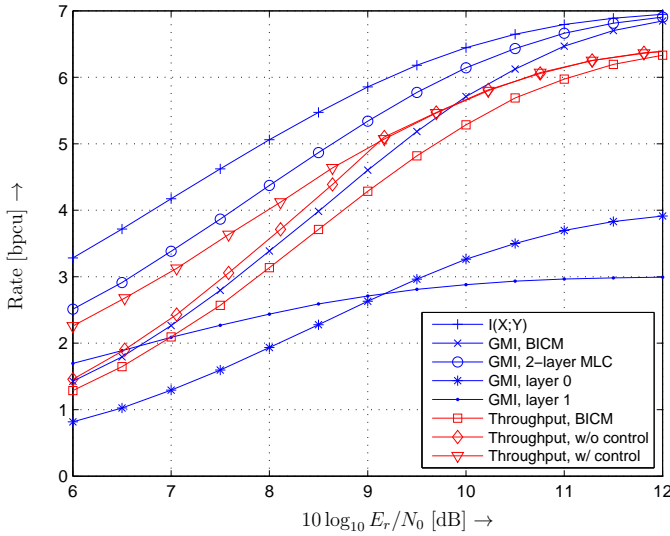
1) *Noncoherent Frequency-Shift Keying (FSK)*: FSK with noncoherent detection is attractive due to its simple detector implementation. In the signal space, each transmit symbol can be represented by a vector  $x = [x_0 \dots x_{M-1}]$  of  $M$  elements where only one element is one and all the others are zero. Let  $h$  be the complex channel gain (assumed to be the same for all frequency bands), and  $n$  be a length- $M$  complex AWGN noise vector with variance  $N_0$  per element. The received signal is a length- $M$  complex vector

$$y = hx + n. \quad (28)$$

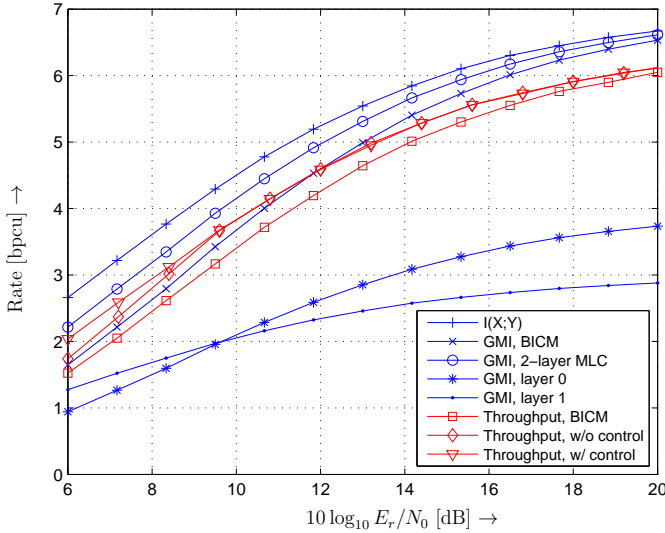
Noncoherent detection requires only knowledge of the magnitude, but not the phase, of the channel gains and received samples. The channel transition probability is given by [12]

$$p_{Y|X}(y|x) \propto I_0\left(\frac{2|h||y_e(x)|}{N_0}\right), \quad (29)$$

where  $I_0(\cdot)$  is the zeroth order modified Bessel function of the first kind, and  $|y_e(x)|$  is the magnitude of the element of  $y$  at the position of the non-zero element of  $x$ . We consider 128-FSK [27] with  $m = 7$  levels,  $\kappa = 2$  layers, and matched decoding metric (27). The lower four levels are grouped into layer 0 and the upper three levels are combined into layer 1. Similarly to the MIMO example, the SNR is defined as  $E_r/N_0$ , and for this modulation  $E_r = \mathcal{E}\{|h|^2\}$ . The layer GMIs are plotted in Figure 7(a) for the case of an AWGN channel and in Figure 7(b) for the case of an i.i.d. Rayleigh fading channel. Depending on the SNR, the GMI of layer 0 can be smaller or larger than that of layer 1. Therefore, the MLC system can be unstable or stable, respectively. We recall that when the system is stable, no control is needed and no structural loss results. For all SNR values, we apply our rateless scheme (i) without control and (ii) with MSL using  $\ell_{\min} = 1.06K/C$ . As expected, Figure 7 shows that when the system is unstable, MSL improves the throughput. Furthermore, when the system is stable, MSL practically does



(a) AWGN channel



(b) i.i.d. Rayleigh fading channel

Fig. 7. GMI and throughput of rateless MLC for 128-FSK. Label “w/o control” and “w/ control” are for the rotational scheme without and with MSL control, respectively. (a) AWGN channel. (b) i.i.d. Rayleigh fading channel.

not incur structural loss, achieving the same throughput as the case without control. This suggests that we can safely apply MSL control to the whole range of SNR values.

Rateless transmission is most beneficial in slow-fading channel environments. Thus, we now illustrate the performance of our scheme in a slow fading example. To this end, consider a channel which evolves from one symbol to the next according to the first-order autoregressive (AR1) model

$$h_i = ah_{i-1} + \sqrt{1-a^2}w_i, \quad (30)$$

where  $w_i$  is an i.i.d. zero mean complex Gaussian process with variance  $E_r$ . The parameter  $0 \leq a \leq 1$  determines the rate of fading. The extreme values  $a = 1$  and  $a = 0$  turn (30) into an AWGN channel and an i.i.d. Rayleigh

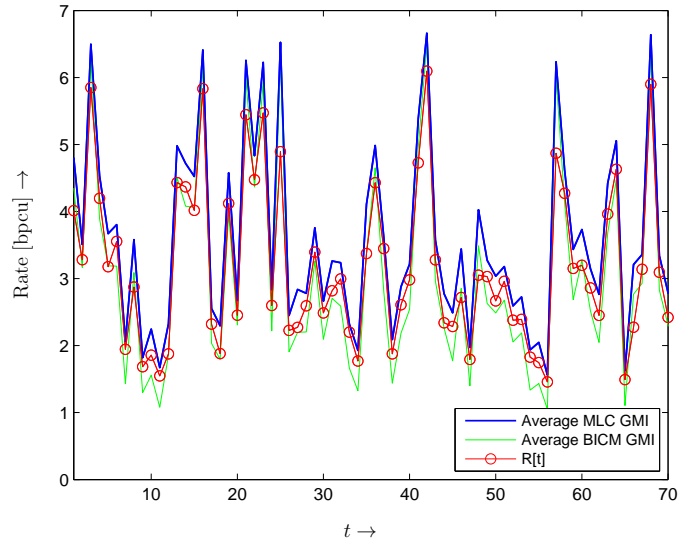


Fig. 8. Example of 2-layer rateless MLC over slow fading channel. The value  $R[t]$  approximates the average throughput during time interval  $t$ .

fading channel, respectively. In slow fading cases, we modify the MSL control to: always receive at least an amount of information  $\mathcal{I}_{\min}$  in each segment. Corresponding to  $N_0$  and the magnitude of the instantaneous channel gain  $|h_i|$ , the instantaneous GMI can be obtained from the GMI plots in Figure 7(a). Let  $C[t]$  be the MLC GMI averaged over time interval  $t$ . We use the value  $R[t] = C[t]K/\theta[t]$  as an indication of the throughput in interval  $t$ . We recall that  $\theta[t]$  (13) is the total amount of information associated with the codeword that has just been decoded in this interval. The ratio  $K/\theta[t]$  is the relative throughput efficiency of the codeword. As an interesting example, we consider  $10 \log_{10}(E_r/N_0) = 8$  dB and  $a = 0.999$  such that  $C[t]$  would change significantly across intervals. We use MSL control with  $\mathcal{I}_{\min} = 1.06K$ . Samples of average MLC GMI and  $R[t]$  obtained from a simulation are plotted in Figure 8. For comparison, the BICM GMI of the same channel realizations, averaged over the same segment boundaries, is also included. We observe that the throughput of our scheme closely follows the associated MLC GMI. Furthermore,  $R[t]$  exceeds the BICM GMI for the majority of the intervals. Averaging over the whole 70 sample intervals illustrated in Figure 8, the MLC GMI is 3.03 bpcu,  $R[t]$  is 2.70 bpcu, and the BICM GMI is 2.48 bpcu. That is, our scheme attains a throughput gain in slow fading channels as well.

2) *Pulse-Position Modulation (PPM)*: PPM is a power-efficient signaling for fiber and free-space optical communications. Similarly to FSK, each  $M$ -ary PPM symbol is a vector of  $M$  elements  $x = [x_0 \dots x_{M-1}]$  with exactly one element of one (the *on* slot) and  $M - 1$  elements of zero (the *off* slots). We consider an FSO transmission scenario for which the memoryless photon counting channel model is applicable [28]. The received sample is a vector  $y = [y_0 \dots y_{M-1}]$  with

$$y_i = x_i s_i + n_i, \quad (31)$$

where  $s_i$  and  $n_i$  are i.i.d. Poisson random variables with mean  $\lambda_s$  and  $\lambda_b$  respectively. Thus,  $\lambda_s$  is the average power of

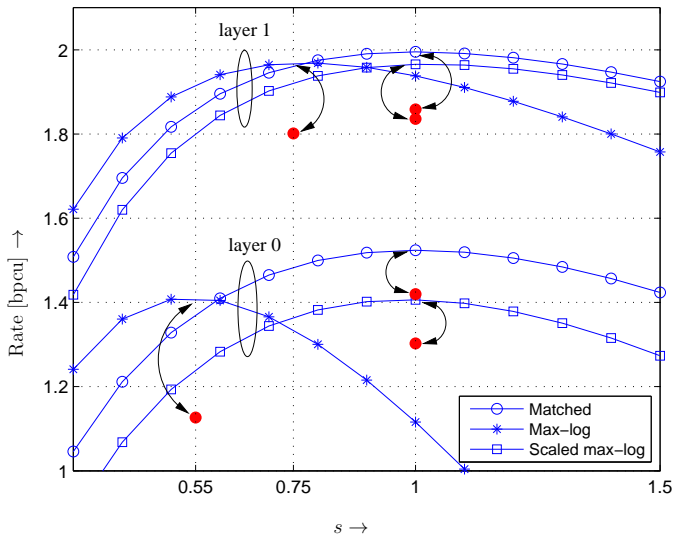


Fig. 9. Layer I-curves of the 128-PPM MLC scheme with  $\lambda_b = 0.2$  photons/slot and  $10 \log_{10}(\lambda_s/(M\lambda_b)) = -10.5$  dB and different metrics. Each doubled-headed arrow connects the peak of an I-curve with a marker that represents the associated throughput with an off-the-shelf Raptor code.

the pulses and  $\lambda_b$  is the average power of the background radiation. The channel transition probability is given by

$$p(y_i|x_i) = \frac{(\lambda_s x_i + \lambda_b)^{y_i}}{y_i!} \exp(-[\lambda_s x_i + \lambda_b]), \quad (32)$$

and  $p(y|x) = \prod_{i=0}^M p(y_i|x_i)$ . Taking into account the fact that all PPM symbols have exactly one element of 1 and  $M-1$  elements of 0, we have

$$p(y|x) \propto \left(1 + \frac{\lambda_s}{\lambda_b}\right)^{y_e(x)} \quad (33)$$

where  $y_e(x)$  is the element of  $y$  at the position of the non-zero element of  $x$ . We consider 128-PPM with background radiation  $\lambda_b = 0.2$  photon/slot and MLC with  $\kappa = 2$  layers where the lower four bits are in layer 0 and the upper three bits are in layer 1. Simulation results with the matched metric (27) demonstrate the same trends as in the case of FSK, and therefore are not presented here. We now consider the popular max-log approximate LLR metric, which is

$$\Lambda_{q_{B_i, Y, \hat{D}_i}}(y, d) = \left( \max_{x \in \mathcal{X}_{d_i, i}^{d_i, 0}} y_e(x) - \max_{x \in \mathcal{X}_{d_i, i}^{d_i, 1}} y_e(x) \right) \ln \left( 1 + \frac{\lambda_s}{\lambda_b} \right). \quad (34)$$

This max-log LLR metric requires less computation than the matched metric (27).

We consider a mid-range SNR of  $10 \log_{10}(\lambda_s/(M\lambda_b)) = -10.5$  dB, at which the 2-layer MLC scheme with matched metric attains a GMI of 3.5 bpcu. Figure 9 shows the I-curves (9) of the two layers with matched and max-log metrics. Due to the equidistant property of the PPM signal points, binary I-curves of all levels within a layer are the same. Thus, the I-curve of layer 0 is equal to four times the binary I-curve of any level in layer 0. Similarly, the I-curve of layer 1 is equal to three times the binary I-curve of any level in that layer. From

Figure 9, we see that the max-log metric incurs only a small reduction in the GMI. The critical point (the  $s$ -coordinate of the peak) of the I-curve of layer 0 is 0.55 and that of layer 1 is 0.75. As the critical point of the I-curve is not equal to 1, the throughput achieved by sum-product SBS decoding can be considerably lower than the GMI [8]. We illustrate this phenomenon via a simulation in which the Raptor decoder is fed with LLR samples from layer 0 or layer 1 alone. In the detection at layer 1, we use transmitted data from layer 0 instead of the estimated data. The resulting throughputs are presented as bold markers in Figure 9. The phenomenon is especially pronounced at layer 0: while the throughput with the matched metric is about 93% of the corresponding GMI, this ratio reduces to only 80% for the max-log metric. We now apply LLR scaling according to (12). That is, we multiply the max-log LLR at layer 0 with 0.55 and at layer 1 with 0.75. Again, we note that the scaling factor for layer 1 is obtained without considering error propagation from layer 0. This scaling shifts the critical point of the I-curve to 1 with the GMI remains the same. The scaled max-log metric results in an improved throughput, as shown in Figure 9. Finally, using the scaled max-log metric in our rotational rateless scheme with appropriate  $\ell_{\min}$ , we obtain a throughput of 3.1 bpcu. In comparison, a simulation of matched BICM with the same Raptor code results in a throughput of 2.5 bpcu. Thus, scaled max-log 2-layer MLC attains a throughput gain of more than 20% compared to matched BICM.

## VI. CONCLUSION

In this paper we considered two practical issues of MLC. Our first contribution is the rate analysis and a simple metric mismatch correction technique for MLC with arbitrary decoding metrics. To this end, we used the GMI as an approximation to the maximum achievable rate of the layer, which can be estimated separately and with the assumption that there are no errors in the decision at lower layers. Our second contribution is the proposal of a novel rateless MLC scheme with no or negligible structural rate loss. We provided extensive numerical examples with MIMO, FSK, and PPM signaling in stationary or slow-fading scenario, using matched or max-log metric, which demonstrate that our scheme can attain throughput gains compared to rateless BICM.

## APPENDIX PROOF OF THEOREM 1

We first show achievability. Let  $C_k$  and  $m_k$  be an achievable rate and the number of levels (bits) in layer  $k$ , respectively. That is, in layer  $k$ , conditioned on correct decoding at layers 0 to  $k-1$ , for any  $\epsilon_k > 0$  and large enough  $N$ , there exists a code  $\mathcal{C}_k(m_k N', R_k)$  of length  $m_k N' > N$  and rate  $R_k \geq C_k$  and a decoding algorithm  $\mathcal{D}_k$  such that the block error probability  $P_k \leq \epsilon_k$ . MLC with these codes has rate  $R \geq C = \sum_{k=0}^{\kappa-1} C_k$ . Furthermore, the block error probability of MLC is upper bounded by [29, Inequality 4.1]

$$P \leq \sum_{k=0}^{\kappa-1} P_k \leq \sum_{k=0}^{\kappa-1} \epsilon_k = \epsilon. \quad (35)$$

Hence, rate  $C$  is achievable.

Now let  $C_k$  be the maximum achievable rate for each layer  $k$  and  $C = \sum_{k=0}^{\kappa-1} C_k$ . Transmission with rate  $R > C$  by the MLC scheme requires that at least one layer, say  $k'$ , has to transmit with rate  $R_{k'} > C_{k'}$ . Hence, there exists an  $\epsilon > 0$  such that for any code and decoding algorithm at this layer, the block error probability is  $P_{k'} > \epsilon$ . Since

$$P \geq P_{k'} \quad \forall k', \quad (36)$$

$C$  is the maximum achievable rate for MLC.

## REFERENCES

- [1] S. Haykin, *Communication Systems*, 4th ed. New York, NY: Wiley, 2000.
- [2] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, pp. 371–377, 1977.
- [3] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1367–1391, Jul. 1999.
- [4] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.
- [5] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, pp. 927–946, May 1998.
- [6] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1-2, pp. 1–135, 2008.
- [7] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, Jun. 2009.
- [8] T. T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.
- [9] C. Stierstorfer and R. Fischer, "(Gray) mappings for bit-interleaved coded modulation," in *IEEE VTC Spring*, Apr. 2007, pp. 1703–1707.
- [10] S. H. Müller-Weinfurtner, "Coding approaches for multiple antenna transmission in fast fading and OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2442–2450, 2002.
- [11] M. McKay and I. Collings, "Capacity and performance of MIMO-BICM with zero-forcing receivers," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 74–83, Jan. 2005.
- [12] J. Proakis, *Digital Communications*, 4th ed. New York, NY: McGraw-Hill, 2001.
- [13] J. Hamkins and B. Moision, "Multipulse pulse-position modulation on discrete memoryless channels," *JPL Interplanetary Network Progress Report*, vol. 42, p. 161, May 2005.
- [14] L. Lampe, R. Schober, and R. Fischer, "Multilevel coding for multiple-antenna transmission," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 203–208, Jan. 2004.
- [15] T. T. Nguyen and L. Lampe, "Coded multipulse pulse-position modulation for free-space optical communications," *IEEE Trans. Commun.*, vol. 58, no. 4, pp. 1036–1041, Apr. 2010.
- [16] J. Jaldén, P. Fertl, and G. Matz, "On the generalized mutual information of BICM systems with approximate demodulation," in *IEEE Inf. Theory Workshop*, Cairo, Egypt, Jan. 2010.
- [17] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Archiv für Elektronik & Übertragungstechnik (AEÜ)*, vol. 47, no. 4, pp. 228–239, 1993.
- [18] N. Merhav, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.
- [19] M. Luby, "LT codes," in *Proc. 43rd IEEE Symp. Found. Comp. Sc. (FOCS)*, Vancouver, BC, Canada, Nov. 2002, pp. 271–280.
- [20] A. Shokrollahi, "Raptor codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2551–2567, Jun. 2006.
- [21] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
- [22] D. J. C. MacKay, *Information Theory Inference and Learning Algorithms*. Cambridge Univ. Press, 2003, version 7.0.
- [23] R. Unbehauen, *Systemtheorie*. Oldenbourg, 1993.
- [24] E. F. Camacho and C. Bordons, *Model Predictive Control*. Springer, Jun. 1999.
- [25] R. Palanki and J. Yedidia, "Rateless codes on noisy channels," 2004. [Online]. Available: <http://www.merl.com/reports/docs/TR2003-124.pdf>
- [26] O. Etesami and A. Shokrollahi, "Raptor codes on binary memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2033–2051, May 2006.
- [27] R. Khalona, "Unified performance analysis of Reed-Solomon coded M-ary FSK modulation in AWGN, Rician and Rayleigh fading channels," in *Intl. Conf. Universal Personal Commun. (ICUPC)*, vol. 2, Oct. 1993, pp. 662–668.
- [28] S. J. Dolinar, J. Hamkins, B. E. Moision, and V. A. Vlnrotter, "Optical modulation and coding," in *Deep Space Optical Communications*, H. Hemmati, Ed. Wiley-Interscience, Apr. 2006.
- [29] U. Wachsmann, *Coded Modulation: Theoretical Concepts and Practical Design Rules*. Aachen, Germany: Shaker Verlag, 1999.



**Trung Thanh Nguyen** (S'06) is a Ph.D. Candidate in the Department of Electrical and Computer Engineering, The University of British Columbia (UBC). He obtained his B.E. in telecommunications in 2004 from the University of Sydney, Australia and his M.A.Sc. in electrical engineering from UBC in 2006. His current research interests are in the area of communications.



**Lutz Lampe** (M'02, SM'08) received the Diplom (Univ.) and the Ph.D. degrees in electrical engineering from the University of Erlangen, Germany, in 1998 and 2002, respectively. Since 2003 he has been with the Department of Electrical and Computer Engineering at the University of British Columbia, where he is a Full Professor.

He is (co-)recipient of a number of Best Paper Awards, including awards at the 2006 IEEE International Conference on Ultra-Wideband (ICUWB), 2010 IEEE International Conference on Communications (ICC), and 2011 IEEE International Conference on Power Line Communications (ISPLC). In 2003, he received the Dissertation Award of the German Society of Information Techniques (ITG). He was awarded the UBC Killam Research Prize in 2008, the Friedrich Wilhelm Bessel Research Award by the Alexander von Humboldt Foundation in 2009, and the UBC Charles A. McDowell Award of Excellence in Research in 2010. He is an Editor for the IEEE Transactions on Wireless Communications and the IEEE Communications Surveys and Tutorials, and he has served as Associate Editor for the IEEE Transactions on Vehicular Technology from 2004 to 2008 and the International Journal on Electronics and Communications (AEUE) from 2007 to 2011. He was General Chair of the 2005 International Symposium on Power Line Communications (ISPLC) and the 2009 IEEE International Conference on Ultra-Wideband (ICUWB). He is Chair of the IEEE Communications Society Technical Committee on Power Line Communication.