



CHALMERS

Chalmers Publication Library

On the Asymptotic Performance of Bit-Wise Decoders for Coded Modulation

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

IEEE Transactions on Information Theory (ISSN: 0018-9448)

Citation for the published paper:

Ivanov, M. ; Alvarado, A. ; Brännström, F. (2014) "On the Asymptotic Performance of Bit-Wise Decoders for Coded Modulation". *IEEE Transactions on Information Theory*, vol. 60(5), pp. 2796-2804.

<http://dx.doi.org/10.1109/TIT.2014.2312726>

Downloaded from: <http://publications.lib.chalmers.se/publication/198521>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

On the Asymptotic Performance of Bit-Wise Decoders for Coded Modulation

Mikhail Ivanov, Alex Alvarado, *Member, IEEE*, Fredrik Brännström, *Member, IEEE*, and Erik Agrell, *Senior Member, IEEE*

Abstract—Two decoder structures for coded modulation over the Gaussian channel are studied: the maximum likelihood symbol-wise decoder, and the (suboptimal) bit-wise decoder based on the bit-interleaved coded modulation paradigm. We consider a 16-ary quadrature amplitude constellation labeled with a Gray labeling. It is shown that the asymptotic loss in terms of pairwise error probability, for any two codewords caused by the bit-wise decoder, is bounded by 1.25 dB. The analysis also shows that the asymptotic loss is zero for a wide range of linear codes, including all rate-1/2 convolutional codes.

Index Terms—Additive white Gaussian noise, Gray code, pairwise error probability, coded modulation, bit-interleaved coded modulation, logarithmic likelihood ratio, pulse-amplitude modulation.

I. INTRODUCTION AND MOTIVATION

CODED MODULATION (CM) is a concatenation of multilevel modulation and a channel code. One popular coded modulation scheme was proposed and analyzed in [1], [2], where convolutional codes (CCs) were used. Due to the trellis structure of the resulting codes, such systems are called trellis-coded modulation (TCM). The TCM decoder finds the codewords at minimum Euclidean distance by exploiting the trellis structure of the code, e.g., by using a symbol-based Viterbi algorithm. Around the same time, multilevel coding (MLC) was presented in [3], where the main idea was to use different binary codes for different bit positions of the constellation points and multiple decoders at the receiver.

Bit-interleaved coded modulation (BICM) is another approach for CM. BICM was initially proposed in [4] and later studied in [5], [6]. In BICM, the encoder and the modulator are separated by a bit-level interleaver. At the receiver side, a suboptimal bit-wise decoder is used, which operates on the L-values provided by the demapper.

It has recently been shown in [7] (see also [8]) that removing the interleaver may improve the performance of BICM over the additive white Gaussian noise (AWGN) channel. Somewhat

This research was supported by the Swedish Research Council, Sweden, under Grant No. 2011-5950, in part by the Ericssons Research Foundation, Sweden, under Grant No. 556016-0680, and in part by the European Community's Seventh's Framework Programme (FP7/2007-2013) under grant agreement No. 271986. The calculations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at C3SE.

M. Ivanov, F. Brännström, and E. Agrell are with the Dept. of Signals and Systems, Chalmers Univ. of Technology, SE-41296 Gothenburg, Sweden (e-mail: {mikhail.ivanov,fredrik.brannstrom,agrell}@chalmers.se).

A. Alvarado is with the Dept. of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom (email: alex.alvarado@ieee.org).

Copyright ©2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

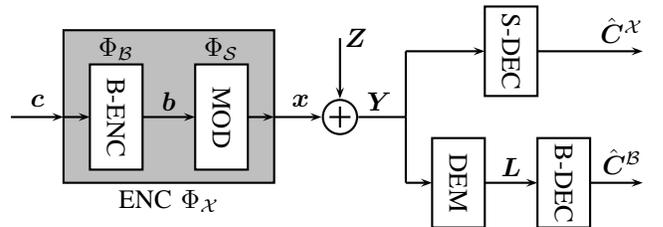


Fig. 1. Block diagram of the analyzed CM system. The CM encoder $\Phi_{\mathcal{X}}$ is used at the transmitter. At the receiver, two decoding algorithms are considered: the ML symbol-wise decoder S-DEC or a suboptimal bit-wise decoder B-DEC.

surprisingly, the results in [8] show that the performance of a bit-wise decoder for an optimized convolutional encoder connected directly to the modulator with a Gray labeling is asymptotically equivalent to the performance of an optimized TCM system. As [9] reveals, these two optimized systems use the same transmitters, i.e., the symbol sequences going into the channel are the same, even though they use different convolutional encoders and binary labelings.

In this paper, we generalize the results in [7], [8] by studying the asymptotic difference between symbol-wise and bit-wise decoders for CM systems with arbitrary binary linear encoders. We consider 16-ary quadrature amplitude modulation (QAM) with a Gray labeling over the AWGN channel. The main result of the paper consists in showing that for any two codewords, the pairwise error probability (PEP) loss caused by the bit-wise decoder is bounded by 1.25 dB. We also prove that for a wide range of linear codes, the asymptotic loss caused by the bit-wise decoder is zero.

II. SYSTEM MODEL

A. Coded Modulation Encoder

Throughout the paper, boldface letters denote vectors or matrices and capital letters denote random variables. The block diagram of the analyzed system is shown in Fig. 1. A CM encoder (ENC) carries out a one-to-one mapping from an information vector of K bits $\mathbf{c} = [c[1], \dots, c[K]] \in \{0, 1\}^K$ to a vector of N symbols $\mathbf{x} = [x[1], \dots, x[N]]$. Each symbol is drawn from a discrete constellation $\mathcal{S} = \{s_1, \dots, s_M\}$, i.e., $x[k] \in \mathcal{S}$ and $k = 1, \dots, N$, where $M = 2^m$ and m is a positive integer. All vectors \mathbf{x} form a CM code $\mathcal{X} \subset \mathcal{S}^N$, where $|\mathcal{X}| = 2^K$ is the number of possible information vectors. The CM encoder is defined as the function $\Phi_{\mathcal{X}} : \{0, 1\}^K \rightarrow \mathcal{X}$ with the corresponding inverse function

$\Phi_{\mathcal{X}}^{-1} : \mathcal{X} \rightarrow \{0, 1\}^K$. Assuming all information vectors to be equally likely, the average energy per symbol can be expressed as $E_s = N^{-1}2^{-K} \sum_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|^2$ and the average energy per bit $E_b = K^{-1}NE_s$.

As all symbols s_i can be uniquely identified by length- m binary labels, any CM encoder described above can be represented as a concatenation of two blocks, as shown in Fig. 1. The modulator (MOD) carries out a one-to-one mapping from m bits to one of the M constellation points. The modulator is defined as the function $\Phi_S : \{0, 1\}^m \rightarrow \mathcal{S}$ with the corresponding inverse function $\Phi_S^{-1} : \mathcal{S} \rightarrow \{0, 1\}^m$. We represent a binary labeling by a vector $\mathbf{q} = [q_1, \dots, q_M]$, where q_i is the integer representation of the m bits mapped to the symbol s_i , with the most significant bit to the left.

A binary encoder (B-ENC) provides the modulator with bits to produce a vector of symbols \mathbf{x} . The B-ENC maps K incoming bits \mathbf{c} into mN coded bits $\mathbf{b} = [\mathbf{b}[1], \dots, \mathbf{b}[N]]$, where $\mathbf{b}[k] = [b_1[k], \dots, b_m[k]] = \Phi_S^{-1}(x[k]) \in \{0, 1\}^m$ and $k = 1, \dots, N$. All vectors \mathbf{b} form a *binary code* $\mathcal{B} \subset \{0, 1\}^{mN}$, where $|\mathcal{B}| = |\mathcal{X}| = 2^K$. The B-ENC is defined as the function $\Phi_B : \{0, 1\}^K \rightarrow \mathcal{B}$ with the corresponding inverse function $\Phi_B^{-1} : \mathcal{B} \rightarrow \{0, 1\}^K$. Throughout the paper, we assume \mathcal{B} to be a binary *linear* code.

The described CM encoder in Fig. 1 generalizes the proposed coding schemes in [2]–[4]. Indeed, it corresponds to TCM if the B-ENC is a terminated convolutional encoder. If the B-ENC is a bank of m parallel encoders, the described encoder represents an MLC encoder. Finally, it corresponds to BICM if the B-ENC includes an interleaver.

When using binary phase-shift keying, the function of the modulator is trivial, and analyzing the CM code \mathcal{X} is equivalent to analyzing a corresponding binary code \mathcal{B} . This, however, is not the case when multilevel modulation is used.

In this paper, we study a 16-QAM constellation labeled with a Gray code, which is used in many wireless standards, see, e.g., [10, Fig. 18-10], [11, Table 7.1.3-1], [12, Fig. 15]. This modulation format can be viewed as a direct product of two Gray labeled 4-ary pulse amplitude modulation (PAM) constellations [13], and therefore, only the constituent 4-PAM constellation needs to be considered. This constellation is defined as $\mathcal{S} = \{-3d, -d, d, 3d\}$, where d is a normalization factor and $s_i < s_{i+1}$.

We consider a real discrete-time memoryless AWGN channel, i.e., given the channel input x , the channel output is $Y = x + Z$, where Z is a zero-mean Gaussian random variable with variance $\sigma_Z^2 = N_0/2$. The conditional probability density function (PDF) of the channel output is

$$p_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma_Z^2}} e^{-\frac{(y-x)^2}{2\sigma_Z^2}}. \quad (1)$$

A Gaussian distribution with mean value μ and variance σ^2 is denoted by $\mathcal{N}(\mu, \sigma^2)$, i.e., $Y \sim \mathcal{N}(x, \sigma_Z^2)$.

It is well known that there are $4! = 24$ labelings for 4-PAM. Due to the symmetry of the constellation and the channel, the labelings $\mathbf{q} = [q_1, q_2, q_3, q_4]$ and $\mathbf{q}' = [q_4, q_3, q_2, q_1]$ will produce equivalent CM codes \mathcal{X} and \mathcal{X}' for any binary code \mathcal{B} , i.e., if a codeword \mathbf{x} belongs to the code \mathcal{X} , then $-\mathbf{x}$ belongs

TABLE I
GRAY LABELINGS FOR 4-PAM

Labeling	\mathbf{q}
GL ₁	[0, 1, 3, 2]
GL ₂	[0, 2, 3, 1]
GL ₃	[1, 0, 2, 3]
GL ₄	[2, 0, 1, 3]

to the code \mathcal{X}' . The number of labelings is therefore reduced to 12. Four of them are Gray labelings, which are listed in Table I. In this paper, only Gray labelings are considered.

The most popular Gray labeling is GL₁, often referred to as the binary reflected Gray code (BRGC) [14]–[16]. All these labelings give the same uncoded bit error rate and BICM generalized mutual information [17] for the AWGN channel, thus, they are usually said to be equivalent [15]. However, in this paper, we consider them separately, as all these labelings produce different CM codes when used with a given binary code \mathcal{B} . The difference between the labelings will be evident later on in Sec. IV.

In this paper, we study two different decoders for the CM encoder in Fig. 1, which we describe below.

B. Symbol-Wise Decoder

The symbol-wise decoder (S-DEC) shown in Fig. 1 performs maximum likelihood (ML) decoding by computing

$$\hat{\mathbf{c}}^{\mathcal{X}} = \Phi_{\mathcal{X}}^{-1} \left(\underset{\mathbf{x} \in \mathcal{X}}{\operatorname{argmin}} \{D^{\mathcal{X}}(\mathbf{x})\} \right), \quad (2)$$

where $D^{\mathcal{X}}(\mathbf{x}) = \sum_{k=1}^N (Y[k] - x[k])^2$. In other words, the S-DEC searches for the closest codeword to the observation $\mathbf{Y} = [Y[1], \dots, Y[N]]$. Assuming the codeword $\mathbf{x} \in \mathcal{X}$ is transmitted, an error occurs if there is a codeword $\hat{\mathbf{x}} = [\hat{x}[1], \dots, \hat{x}[N]] \in \mathcal{X}$, such that $D^{\mathcal{X}}(\mathbf{x}) > D^{\mathcal{X}}(\hat{\mathbf{x}})$. The probability of such an event is called the PEP and can be calculated as

$$\text{PEP}^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) = \Pr\{\Delta^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) < 0\}, \quad (3)$$

where $\Pr\{\cdot\}$ stands for probability and $\Delta^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) \triangleq D^{\mathcal{X}}(\hat{\mathbf{x}}) - D^{\mathcal{X}}(\mathbf{x})$. For future use, we express $\Delta^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}})$ as

$$\Delta^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) = 4d \sum_{k=1}^N \Lambda^{\mathcal{X}}(x[k], \hat{x}[k]), \quad (4)$$

where

$$\begin{aligned} \Lambda^{\mathcal{X}}(x[k], \hat{x}[k]) &\triangleq \frac{(Y[k] - \hat{x}[k])^2 - (Y[k] - x[k])^2}{4d} \\ &= \frac{x[k] - \hat{x}[k]}{2d} Y[k] + \frac{\hat{x}^2[k] - x^2[k]}{4d} \end{aligned} \quad (5)$$

is called a symbol metric difference (SMD).

C. Bit-Wise Decoder

The bit-wise decoder (B-DEC) shown in Fig. 1 operates on the bit reliability metrics provided by a demapper (DEM). The demapper acts independently of the B-DEC and

calculates a vector $\mathbf{L} = [L[1], \dots, L[N]]$, where $L[k] = [L_1[k], \dots, L_m[k]]$ are the logarithmic-likelihood ratios (L-values). We use the so-called max-log approximation [4, eq. (3.2)], [5, eq. (2.15)], [17, eq. (12)] for the calculation of the L-values, i.e.,

$$L_j[k] = \frac{1}{2\sigma_Z^2} \left[\min_{s \in \mathcal{S}_{j,0}} (Y[k] - s)^2 - \min_{s \in \mathcal{S}_{j,1}} (Y[k] - s)^2 \right] \quad (6)$$

with $j = 1, \dots, m$, where $\mathcal{S}_{j,u} \subset \mathcal{S}$ is the subset of constellation points whose labels have the value $u \in \{0, 1\}$ in the j th bit position.

The calculated L-values are passed to the B-DEC, which uses the decoding rule [5, Sec. 2.2], [17, eq. (13)]

$$\hat{\mathbf{C}}^{\mathcal{B}} = \Phi_{\mathcal{B}}^{-1} \left(\underset{\mathbf{b} \in \mathcal{B}}{\operatorname{argmax}} \{ D^{\mathcal{B}}(\mathbf{b}) \} \right), \quad (7)$$

where $D^{\mathcal{B}}(\mathbf{b}) = (2\mathbf{b} - 1)\mathbf{L}^T = \sum_{k=1}^N (2b[k] - 1)\mathbf{L}^T[k]$ and $(\cdot)^T$ denotes transposition.

The PEP for the B-DEC is given by

$$\operatorname{PEP}^{\mathcal{B}}(\mathbf{b}, \hat{\mathbf{b}}) = \Pr\{\Delta^{\mathcal{B}}(\mathbf{b}, \hat{\mathbf{b}}) < 0\}, \quad (8)$$

where $\Delta^{\mathcal{B}}(\mathbf{b}, \hat{\mathbf{b}}) \triangleq D^{\mathcal{B}}(\mathbf{b}) - D^{\mathcal{B}}(\hat{\mathbf{b}})$ is the difference between the metrics for the transmitted codeword \mathbf{b} and the competing codeword $\hat{\mathbf{b}} \in \mathcal{B}$. Since the mapping between \mathbf{b} and \mathbf{x} is one-to-one, with a slight abuse of notation, $\Delta^{\mathcal{B}}(\mathbf{b}, \hat{\mathbf{b}})$ can be written as a function of codewords \mathbf{x} and $\hat{\mathbf{x}}$ instead, i.e.,

$$\Delta^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\sigma_Z^2}{2d} \sum_{k=1}^N \Lambda^{\mathcal{B}}(x[k], \hat{x}[k]), \quad (9)$$

where the SMD in this case is

$$\Lambda^{\mathcal{B}}(x[k], \hat{x}[k]) \triangleq \frac{\sigma_Z^2}{d} (\Phi_{\mathcal{S}}^{-1}(x[k]) - \Phi_{\mathcal{S}}^{-1}(\hat{x}[k]))\mathbf{L}[k]^T. \quad (10)$$

The B-DEC described above corresponds to the standard (noniterative) BICM decoder. We refrain from using this name, as the interleaver might or might not be included in the transmitter. Moreover, if there is an interleaver, we assume it to be part of the B-ENC.

The performance of the B-DEC is highly dependent on the distributions of the L-values. Since the distribution of the L-values may depend on whether zero or one was transmitted, symmetrization techniques are usually used in order to simplify the analysis. In [6, Sec. IV], a time-varying labeling was proposed in order to symmetrize the channel. The same effect can be achieved by using a random scrambler as in [5, Sec. 4.1] or [8, Sec. II]. This allows to simplify the analysis by assuming that the all-zero codeword is transmitted. However, the symmetrization makes the comparison of the two decoders unfair and we do not include a scrambler in our analysis, hence, we consider all possible pairs of codewords when analyzing codes.

III. SYMBOL VS. BIT DECODER

A. Distribution of the SMDs

To compare the PEP for the S-DEC in (3) and the B-DEC in (8), we analyze the distributions of the SMDs in (5) and (10).

TABLE II

DISTRIBUTION PARAMETERS (μ, σ^2) FOR THE SMD (5) OF THE S-DEC. CIRCLES, STARS, AND DIAMONDS SHOW THE ERROR VECTOR \mathbf{e} EQUAL TO $[0, 1]$, $[1, 1]$, AND $[1, 0]$, RESPECTIVELY, FOR GL₃.

$x[k] \backslash \hat{x}[k]$	s_1	s_2	s_3	s_4
s_1	–	$(1, 1)^\circ$	$(4, 4)^*$	$(9, 9)^\diamond$
s_2	$(1, 1)^\circ$	–	$(1, 1)^\circ$	$(4, 4)^*$
s_3	$(4, 4)^*$	$(1, 1)^\circ$	–	$(1, 1)^\circ$
s_4	$(9, 9)^\diamond$	$(4, 4)^*$	$(1, 1)^\circ$	–

TABLE III

DISTRIBUTION PARAMETERS (μ, σ^2) FOR THE SMD (10) OF THE B-DEC. CIRCLES, STARS, AND DIAMONDS SHOW THE ERROR VECTOR \mathbf{e} EQUAL TO $[0, 1]$, $[1, 1]$, AND $[1, 0]$, RESPECTIVELY, FOR GL₃.

$x[k] \backslash \hat{x}[k]$	s_1	s_2	s_3	s_4
s_1	–	$(1, 1)^\circ$	$(4, 4)^*$	$(3, 1)^\diamond$
s_2	$(1, 1)^\circ$	–	$(1, 1)^\circ$	$(4, 4)^*$
s_3	$(4, 4)^*$	$(1, 1)^\circ$	–	$(1, 1)^\circ$
s_4	$(3, 1)^\diamond$	$(4, 4)^*$	$(1, 1)^\circ$	–

Lemma 1: For 4-PAM with any labeling, the SMDs in (5) divided by $4d$ are distributed as

$$\Lambda^{\mathcal{X}}(x[k], \hat{x}[k]) \sim \mathcal{N}(\mu d, \sigma^2 \sigma_Z^2), \quad (11)$$

where (μ, σ^2) are shown in Table II.

Proof: Since the SMDs in (5) are linear functions of the observation $Y[k]$, the SMDs follow a Gaussian distribution. When $x[k] = s_i$ and $\hat{x}[k] = s_j$, the mean value of the SMD is $\mu = (4d^2)^{-1}(2(s_i - s_j)s_i + (s_j^2 - s_i^2)) = (4d^2)^{-1}(s_i - s_j)^2$. The variance can be calculated as $\sigma^2 = (4d^2)^{-1}(s_i - s_j)^2$. Substituting values of s_i and s_j gives the parameters shown in Table II. \square

We note that the results in Lemma 1 are valid for any labeling, not only Gray labelings.

Lemma 2: For 4-PAM with any Gray labeling, the distribution of the SMDs in (10) can be approximated as

$$\Lambda^{\mathcal{B}}(x[k], \hat{x}[k]) \sim \mathcal{N}(\mu d, \sigma^2 \sigma_Z^2), \quad (12)$$

where (μ, σ^2) are shown in Table III.

Proof: Since the L-value in (6) is a piece-wise linear function of the observation, the distribution of the L-value is a superposition of piece-wise Gaussian distributions, with mean and variance defined by the linear pieces and the transmitted symbol. In [18, Sec. 5], [19, Sec. III-C], it has been shown that at high signal-to-noise ratios (SNR), measured as E_s/N_0 or E_b/N_0 , the so-called zero-crossing (ZcMod) approximation of such a PDF gives good results in terms of coded bit-error rate (BER) and mutual information. The results shown in Table III are obtained from [8, Table II] by scaling the SMDs by σ_Z^2 . The distributions are independent of a particular Gray labeling and depend only on the compared symbols. The tightness of the ZcMod approximation will be discussed in Sec. III-C. \square

Comparing Tables II and III, we note that the tables are identical, except for the corner entries in gray. We will use this simple observation in the following section to bound the loss incurred by the B-DEC when compared to the S-DEC.

B. Pairwise Error Probability Analysis

In this section, we study the asymptotic performance of the S-DEC and the B-DEC. Throughout the section, we use GL_3 for illustration, i.e., symbols s_k , $k = 1, \dots, 4$ are labeled with $[0, 1]$, $[0, 0]$, $[1, 0]$, and $[1, 1]$, respectively. All discussions and derivations below apply directly to GL_1 , and also to GL_2 and GL_4 if the labels $[1, 0]$ and $[0, 1]$ are swapped.

Examining Tables II and III, we see that, in many cases, the distribution of the SMDs depends on the binary vector $e \triangleq \Phi_S^{-1}(x[k]) \oplus \Phi_S^{-1}(\hat{x}[k]) \in \{0, 1\}^2$, where \oplus denotes modulo-2 addition. When $e = [0, 0]$, the distributions are not defined (main diagonal of the tables). For $e = [1, 1]$, the distribution parameters are (4, 4) (marked with stars in the tables) and for $e = [0, 1]$, the distribution parameters are (1, 1) (marked with circles). However, the distribution parameters for $e = [1, 0]$ are different (marked with diamonds in the tables). When the compared symbols are s_2 and s_3 , the distribution parameters are (1, 1), whereas the distribution parameters are (9, 9) and (3, 1) for the S-DEC and the B-DEC, respectively, when the compared symbols are s_1 and s_4 (gray entries of the tables). We use $(\mu_{[0,1]}, \sigma_{[0,1]}^2)$ for entries marked with circles, $(\mu_{[1,1]}, \sigma_{[1,1]}^2)$ for entries marked with stars, $(\mu_{[1,0]}, \sigma_{[1,0]}^2)$ for white entries marked with diamonds, and $(\mu_{\mathcal{X}}, \sigma_{\mathcal{X}}^2)$ and $(\mu_{\mathcal{B}}, \sigma_{\mathcal{B}}^2)$ for gray entries marked with diamonds for the S-DEC and the B-DEC, respectively.

We define the set of possible non-zero vectors e as $\mathcal{E} = \{[0, 1], [1, 0], [1, 1]\}$. For two codewords x and \hat{x} and for $e \in \mathcal{E}$, we define $w_e(x, \hat{x})$ as

$$w_e(x, \hat{x}) = \sum_{k=1}^N \mathbb{I}\{\Phi_S^{-1}(x[k]) \oplus \Phi_S^{-1}(\hat{x}[k]) = e\}, \quad (13)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. In other words, $w_e(x, \hat{x})$ is the number of pairs $(x[k], \hat{x}[k])$ in x and \hat{x} such that $\Phi_S^{-1}(x[k]) \oplus \Phi_S^{-1}(\hat{x}[k]) = e$. In addition, we define $w_c(x, \hat{x})$ as the number of pairs $(x[k], \hat{x}[k])$ in x and \hat{x} such that $(x[k], \hat{x}[k]) = (s_1, s_4)$ or $(x[k], \hat{x}[k]) = (s_4, s_1)$, i.e., $w_c(x, \hat{x})$ is the number of *corner* entries (gray entries in Tables II and III). Clearly, $w_{[1,0]}(x, \hat{x}) \geq w_c(x, \hat{x})$, as the former includes pairs of symbols counted in the latter. To simplify the notation, the arguments of $w_e(x, \hat{x})$ and $w_c(x, \hat{x})$ are omitted when the arguments are clearly stated in the text.

From Lemmas 1 and 2, it follows that the SMDs are independent Gaussian random variables. Using the introduced notation, the PEP for the S-DEC and the B-DEC in (3) and (8) can therefore be expressed as

$$\text{PEP}(x, \hat{x}) = Q\left(a(x, \hat{x}) \frac{d}{\sigma_Z}\right), \quad (14)$$

where $Q(\cdot)$ is the Gaussian Q-function and the normalized distance $a(x, \hat{x})$ is either

$$a^{\mathcal{X}}(x, \hat{x}) = \frac{w_c(\mu_{\mathcal{X}} - \mu_{[1,0]}) + \sum_{e \in \mathcal{E}} w_e \mu_e}{\sqrt{w_c(\sigma_{\mathcal{X}}^2 - \sigma_{[1,0]}^2) + \sum_{e \in \mathcal{E}} w_e \sigma_e^2}} \quad (15)$$

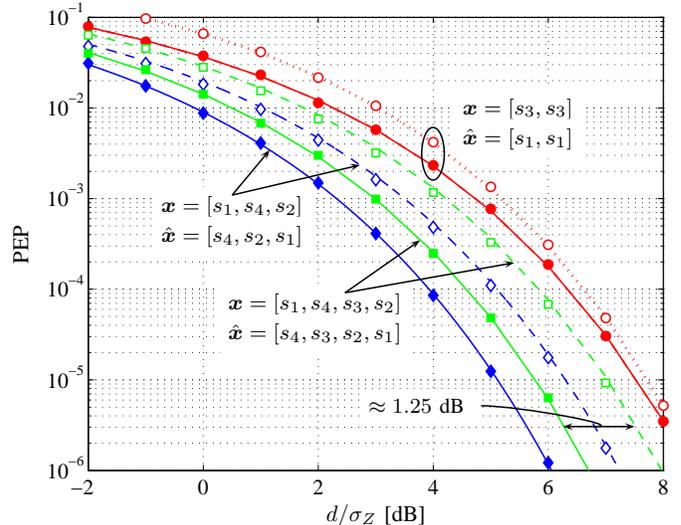


Fig. 2. The PEP for three different pairs of codewords x and \hat{x} . Solid and dashed lines represent analytical PEP in (14) for the S-DEC and the B-DEC, resp. Filled and empty markers show simulation results for the S-DEC and the B-DEC, resp. The dotted line shows the exact PEP for the B-DEC (see Sec. III-C).

for the S-DEC or

$$a^{\mathcal{B}}(x, \hat{x}) = \frac{w_c(\mu_{\mathcal{B}} - \mu_{[1,0]}) + \sum_{e \in \mathcal{E}} w_e \mu_e}{\sqrt{w_c(\sigma_{\mathcal{B}}^2 - \sigma_{[1,0]}^2) + \sum_{e \in \mathcal{E}} w_e \sigma_e^2}} \quad (16)$$

for the B-DEC.

Fig. 2 shows the analytical and the simulated PEP for the S-DEC and the B-DEC as functions of d/σ_Z for three different pairs of codewords x and \hat{x} . We note that d^2/σ_Z^2 is proportional to the SNR. Solid and dashed lines represent analytical PEP in (14) for the S-DEC and the B-DEC, respectively. For the codewords $x = [s_3, s_3]$ and $\hat{x} = [s_1, s_1]$ (circles), the dashed line coincides with the solid line. Filled markers represent simulation results for the S-DEC and are exactly on top of the corresponding solid lines. Empty markers show simulation results for the B-DEC. Empty squares and diamond agree well with the analytically predicted PEP; however, empty circles deviate significantly from the analytical prediction (which is based on the ZcMod approximation). We note that instead, empty circles agree well with the dotted line, which is briefly discussed in the next section.

C. Zero-Crossing Approximation

The SMD in (10) is a linear combination of L-values which depends on the compared symbols and their binary labels. For example, if $x[k] = s_3$ and $\hat{x}[k] = s_1$ and GL_3 is used, the SMD is given by $\Lambda^{\mathcal{B}}(x[k], \hat{x}[k]) = \frac{\sigma_Z^2}{d} (L_1[k] - L_2[k])$. Since the L-values $L_1[k]$ and $L_2[k]$ are piece-wise linear functions of the observation $Y[k]$ [19, Fig. 3], so is the SMD. Thus, its exact distribution is a superposition of piece-wise Gaussian functions. Let λ be the realization of the random variable $\Lambda^{\mathcal{B}}(x[k], \hat{x}[k])$ for a given channel realization $Y[k] = y$. The solid line in Fig. 3 shows the SMD as a function of the channel realization y and the dash-dotted line shows the distribution of the observation $Y[k]$ given the transmitted symbol s_3 for

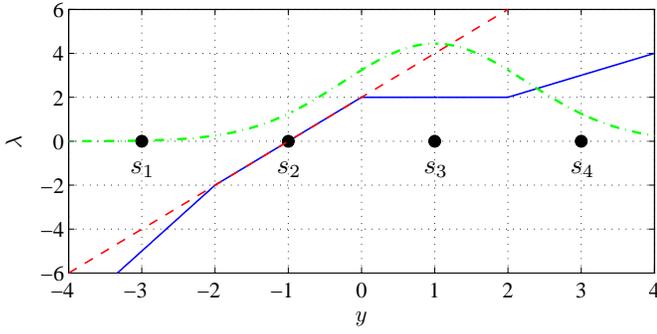


Fig. 3. The SMD $\Lambda^{\mathcal{B}}(s_3, s_1)$ as a function of the observation y (solid line) and the ZcMod approximation of the SMD (dashed line) for $d = 1$. The dash-dotted line shows the distribution $p_{Y|X}(y|s_3)$ (not to scale) for $d/\sigma_Z = -5$ dB.

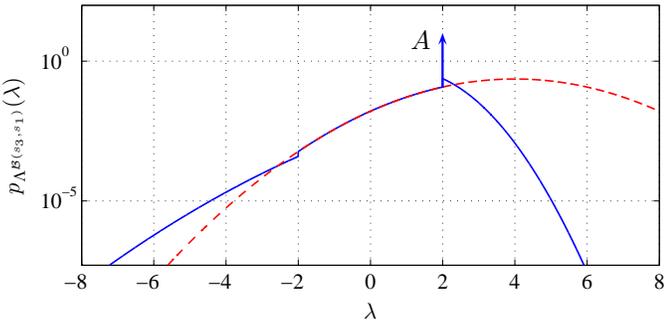


Fig. 4. Distribution of the SMD $\Lambda^{\mathcal{B}}(s_3, s_1)$ for $d/\sigma_Z = -5$ dB. The solid line is the exact PDF and the dashed line shows the approximated PDF using the ZcMod approximation.

$d/\sigma = -5$ dB. The corresponding exact distribution of the SMD $p_{\Lambda^{\mathcal{B}}(s_3, s_1)}(\lambda)$ is shown in Fig. 4 with a solid line and contains a delta function with amplitude $A = 1 - 2Q(d/\sigma_Z)$ due to the horizontal piece of the SMD function in Fig. 3.

The exact PDF is difficult to analyze and approximations are usually used. The ZcMod approximation approximates the SMD with a straight line. The line is chosen so that the exact PDF of the SMD is accurately approximated around $\lambda = 0$. Therefore the line is chosen as a tangent to the SMD function at $\lambda = 0$, i.e., at the zero-crossing. If the SMD function has multiple zero-crossings, then the zero-crossing closest to the transmitted symbol is used. The ZcMod approximation of the SMD $\Lambda^{\mathcal{B}}(s_3, s_1)$ is shown with a dashed line in Fig. 3 and it results in a Gaussian distribution shown with a dashed line in Fig. 4. As can be seen from Fig. 4, the ZcMod approximation approximates well the exact PDF around $\lambda = 0$, which makes the ZcMod approximation suitable for the analysis.

Although the ZcMod approximation has been shown to be good in terms of coded bit-error rate (BER) and mutual information [18, Sec. 5], [19, Sec. III-C], a rigorous proof of its tightness is still missing. This is mainly because such a proof would require to consider all possible pairs of codewords. In the following, we show that the approximation is asymptotically tight for the codewords $\mathbf{x} = [s_3, s_3]$ and $\hat{\mathbf{x}} = [s_1, s_1]$ (circles in Fig. 2).

For the codewords $\mathbf{x} = [s_3, s_3]$ and $\hat{\mathbf{x}} = [s_1, s_1]$, $\Delta^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}})$

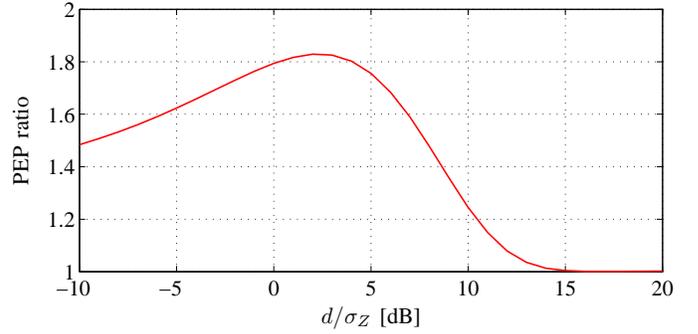


Fig. 5. The ratio between the exact PEP (obtained numerically) using the exact PDF of the SMDs and the PEP predicted by the ZcMod approximation for the codewords $\mathbf{x} = [s_3, s_3]$ and $\hat{\mathbf{x}} = [s_1, s_1]$ (circles in Fig. 2).

in (9) is a sum of two SMDs with the distributions in Fig. 4. When calculating the PEP, a convolution of these PDFs needs to be calculated. The peculiarity of these SMDs, when s_3 is transmitted and s_1 is a competitor, is that the exact PDFs contain a Dirac delta function. When two such PDFs are convolved, the resulting PEP is not well approximated by the ZcMod approximation. The exact PEP calculated numerically using the exact PDF is shown with a dotted line in Fig. 2 and, as expected, it coincides with the simulations for the B-DEC (empty circles).

To study the asymptotic tightness of the ZcMod approximation, we show in Fig. 5 the ratio between the exact PEP and the approximated PEP. This figure shows that for moderate SNR, the ZcMod approximation underestimates the PEP. However, the approximation is tight when $d/\sigma_Z \rightarrow \infty$. This result was also verified analytically by considering upper and lower bounds on the exact PEP. Analogous results were obtained for other pairs of codewords. A behavior similar to what is shown in Fig. 5 will be observed later on in Sec. IV-C.

D. Asymptotic Pairwise Loss

Using (14) and (15)–(16), we define the asymptotic loss (when $d/\sigma_Z \rightarrow \infty$) caused by the B-DEC (compared to the S-DEC) for any two pairs of codewords \mathbf{x} and $\hat{\mathbf{x}}$ as

$$L(\mathbf{x}, \hat{\mathbf{x}}) \triangleq 20 \log_{10} \left(\frac{a^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}})}{a^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}})} \right). \quad (17)$$

The following theorem gives a bound on (17).

Theorem 1: For 4-PAM with any Gray labeling, $L(\mathbf{x}, \hat{\mathbf{x}}) \leq 1.25$ dB for any two codewords \mathbf{x} and $\hat{\mathbf{x}}$.

Proof: Substituting the values in Tables II and III into (15)–(16), the normalized distances can be expressed as

$$a^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\beta + 8w_c}, \quad (18)$$

$$a^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}}) = \beta^{-1/2}(\beta + 2w_c), \quad (19)$$

where

$$\beta = \sum_{\mathbf{e}} w_{\mathbf{e}} \mu_{\mathbf{e}} = \sum_{\mathbf{e}} w_{\mathbf{e}} \sigma_{\mathbf{e}}^2. \quad (20)$$

The loss in (17) is then given by

$$L(\mathbf{x}, \hat{\mathbf{x}}) = 20 \log_{10} \left(\frac{\sqrt{\beta(\beta + 8w_c)}}{\beta + 2w_c} \right). \quad (21)$$

The argument of the logarithm in (21) is a positive function of β and w_c with a single maximum at $\beta = 4w_c$. The maximum value is $\frac{2}{\sqrt{3}}$, which gives $L(\mathbf{x}, \hat{\mathbf{x}}) \leq 1.25$ dB. \square

From the proof of Theorem 1 it follows that the loss is zero if $w_c = 0$ and it achieves its maximum if β in (20) is equal to $4w_c$. Using Tables II and III, it is easy to show that the latter condition is fulfilled for the pair of codewords $\mathbf{x} = [s_1, s_4, s_3, s_2]$ and $\hat{\mathbf{x}} = [s_4, s_3, s_2, s_1]$, and the asymptotic loss is 1.25 dB, as illustrated by the simulation and analytical results (squares) in Fig. 2.

IV. ASYMPTOTIC LOSS FOR CODES

When all the codewords of a code are considered (e.g., in a union bound-type of expression [20, Ch. 4]), only the pairs of codewords at minimum distance will define the high-SNR performance. Since there is no scrambler in the system, the analysis cannot be based on the assumption that the all-zero codeword is transmitted, i.e., all pairs of codewords need to be considered. Hence, the asymptotic loss for a given code \mathcal{B} has to be defined as

$$L(\mathcal{B}) \triangleq 20 \log_{10} \left(\frac{\min_{\mathbf{x} \neq \hat{\mathbf{x}}} a^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}})}{\min_{\mathbf{x} \neq \hat{\mathbf{x}}} a^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}})} \right). \quad (22)$$

In this section, we study the asymptotic loss in (22). We first consider an arbitrary linear code and then discuss a particular case of rate-1/2 CCs.

A. Any Linear Code

The next corollary is a straightforward implication of Theorem 1.

Corollary 1: For 4-PAM with any Gray labeling and any linear code, $L(\mathcal{B}) \leq 1.25$ dB. There exist CM codes for which this bound is exact.

Proof: Let codewords $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ minimize the denominator of (22), i.e., $a^{\mathcal{B}}(\mathbf{x}_i, \mathbf{x}_j) = \min_{\mathbf{x} \neq \hat{\mathbf{x}}} a^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}})$. The asymptotic loss for the code can then be written as

$$\begin{aligned} L(\mathcal{B}) &= 20 \log_{10} \left(\frac{\min_{\mathbf{x} \neq \hat{\mathbf{x}}} a^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}})}{a^{\mathcal{B}}(\mathbf{x}_i, \mathbf{x}_j)} \right) \\ &\leq 20 \log_{10} \left(\frac{a^{\mathcal{X}}(\mathbf{x}_i, \mathbf{x}_j)}{a^{\mathcal{B}}(\mathbf{x}_i, \mathbf{x}_j)} \right), \end{aligned}$$

which proves the first part. To prove the second part, we give an example of such a code. Consider a linear code consisting of two codewords $\mathbf{b}_1 = [0, 0, 0, 0, 0, 0, 0, 0]$ and $\mathbf{b}_2 = [1, 0, 0, 1, 0, 1, 1, 1]$ used with 4-PAM and GL₁. This corresponds to a CM code with two codewords $\mathbf{x}_1 = [s_1, s_1, s_1, s_1]$ and $\mathbf{x}_2 = [s_4, s_2, s_2, s_3]$. From Tables II and III, it follows that for these two codewords $\beta = 4w_c$. Hence, $L(\mathcal{B}) = L(\mathbf{x}_1, \mathbf{x}_2) = 1.25$ dB. \square

Even though linear codes with nonzero asymptotic loss exist, they are not very common due to their special structure, i.e., the closest paths should consist of a special combination of symbols. In what follows, we show that for some labelings and a wide range of linear codes, $w_c = 0$ for the codewords at minimum distance, and therefore, the asymptotic loss in (22) is zero.

Theorem 2: For 4-PAM with GL₃ or GL₄ and any linear code, the loss $L(\mathcal{B}) = 0$.

Proof: Consider the GL₃ labeling. Let \mathbf{x} and $\hat{\mathbf{x}}$ be two different codewords of the code \mathcal{X} with corresponding binary codewords $\mathbf{b}, \hat{\mathbf{b}} \in \mathcal{B}$, such that $w_c(\mathbf{x}, \hat{\mathbf{x}}) \neq 0$. For any linear code, $\mathbf{b}' = \mathbf{b} \oplus \mathbf{b} = [0, \dots, 0]$ and $\hat{\mathbf{b}}' = \hat{\mathbf{b}} \oplus \mathbf{b}$ are also codewords of \mathcal{B} with corresponding $\mathbf{x}', \hat{\mathbf{x}}' \in \mathcal{X}$. As $\mathbf{b}' \oplus \hat{\mathbf{b}}' = \mathbf{b} \oplus \hat{\mathbf{b}}$, we conclude that $w_e(\mathbf{x}', \hat{\mathbf{x}}') = w_e(\mathbf{x}, \hat{\mathbf{x}})$, $\forall e \in \mathcal{E}$. From Tables II and III, it is clear that $w_c(\mathbf{x}', \hat{\mathbf{x}}') = 0$, as $\mathbf{x}' = [s_2, s_2, \dots, s_2]$. Using (18) and the assumption that $w_c(\mathbf{x}, \hat{\mathbf{x}}) \neq 0$, we conclude that for the S-DEC

$$a^{\mathcal{X}}(\mathbf{x}, \hat{\mathbf{x}}) = \sqrt{\beta + 8w_c} > \sqrt{\beta} = a^{\mathcal{X}}(\mathbf{x}', \hat{\mathbf{x}}').$$

Using (19) we show, in a similar way, that for the B-DEC

$$a^{\mathcal{B}}(\mathbf{x}, \hat{\mathbf{x}}) = \beta^{-1/2}(\beta + 2w_c) > \sqrt{\beta} = a^{\mathcal{B}}(\mathbf{x}', \hat{\mathbf{x}}').$$

We showed that $a(\mathbf{x}', \hat{\mathbf{x}}') < a(\mathbf{x}, \hat{\mathbf{x}})$ for both the S-DEC and the B-DEC. Hence, for any two codewords \mathbf{x} and $\hat{\mathbf{x}}$ with $w_c(\mathbf{x}, \hat{\mathbf{x}}) \neq 0$, there always exist two other codewords \mathbf{x}' and $\hat{\mathbf{x}}'$ with $w_c(\mathbf{x}', \hat{\mathbf{x}}') = 0$ at a smaller distance. The latter means that $w_c(\mathbf{x}, \hat{\mathbf{x}}) = 0$ for any pair of codewords \mathbf{x} and $\hat{\mathbf{x}}$ at minimum distance, and hence, the loss in (22) is zero. Similar reasoning directly applies to GL₄. This completes the proof. \square

Remark 1: In other words, the proof of Theorem 2 shows that for 4-PAM with GL₃ or GL₄ and a linear code, among all pairs of codewords $\mathbf{x} \neq \hat{\mathbf{x}}$ at minimum distance, there is one pair such that one of the codewords corresponds to the all-zero binary codeword. For this case, the minimum distance analysis can assume that the all-zero codeword was transmitted.

The peculiar property of GL₃ and GL₄ is that the all-zero label is assigned to one of the innermost constellation points, which guarantees that $\mathbf{x} = [s_2, s_2, \dots, s_2] \in \mathcal{X}$. This is not the case for the GL₁ and GL₂ labelings, where the all-zero label is assigned to one of the outermost symbols. However, for these labelings it is still possible to define a family of codes for which the loss is also zero. This is done in the following theorem.

Theorem 3: For 4-PAM with GL₁, the loss $L(\mathcal{B}) = 0$ if the linear code \mathcal{B} contains a codeword $\mathbf{b}'' = [b''[1], \dots, b''[N]] \in \mathcal{B}$, such that $b''_2[k] = 1, \forall k$. Similarly, for 4-PAM with GL₂, $L(\mathcal{B}) = 0$ if $\mathbf{b}'' \in \mathcal{B}$ and $b''_1[k] = 1, \forall k$.

Proof: First, we assume that GL₁ is used and a codeword \mathbf{b}'' , such that $b''_2[k] = 1, \forall k$, belongs to the code \mathcal{B} . Let \mathbf{x} and $\hat{\mathbf{x}}$ be codewords of the code \mathcal{X} with corresponding binary codewords $\mathbf{b}, \hat{\mathbf{b}} \in \mathcal{B}$, such that $w_c(\mathbf{x}, \hat{\mathbf{x}}) \neq 0$. For a linear code, $\mathbf{b}' = \mathbf{b} \oplus \mathbf{b} \oplus \mathbf{b}''$ and $\hat{\mathbf{b}}' = \hat{\mathbf{b}} \oplus \mathbf{b} \oplus \mathbf{b}''$ are also codewords of \mathcal{B} with corresponding $\mathbf{x}', \hat{\mathbf{x}}' \in \mathcal{X}$. From Tables II and III, it is clear that $w_c(\mathbf{x}', \hat{\mathbf{x}}') = 0$, as $\mathbf{x}' = [x'[1], \dots, x'[N]]$, where $x'[k] \in \{s_2, s_3\}, \forall k$. The rest of the proof is similar to the proof of Theorem 2. Swapping the first and the second bit positions in GL₁, we can analogously prove the second statement for GL₂. \square

B. Rate-1/2 Convolutional Codes

Bringing together the results for different labelings (Theorems 2 and 3), the conclusion is as follows.

Corollary 2: For 4-PAM with any Gray labeling, $L(\mathcal{B}) = 0$ if the linear code \mathcal{B} contains codewords $\mathbf{b}'', \mathbf{b}''' \in \mathcal{B}$, such that $b_1''[k] = 1, \forall k$ and $b_2'''[k] = 1, \forall k$.

Many codes satisfy the conditions in Corollary 2, for instance, all extended Hamming codes, all Reed-Muller codes, all extended BCH codes, and all extended Golay codes. All these codes include the all-one codeword. The codes are extended as they should be of an even length to match the constellation. For such codes, all the four Gray labelings are equivalent, in the sense that for a given binary code they produce four *different* CM codes, with the *same* minimum distance for both the S-DEC and the B-DEC.

Rate-1/2 CCs are of particular interest, as they allow an easy implementation of the ML decoder based on the Viterbi algorithm. In the following theorem, we show that all rate-1/2 CCs also give a zero asymptotic loss.

Theorem 4: For 4-PAM with any Gray labeling and any rate-1/2 CC, $L(\mathcal{B}) = 0$.

Proof: Any rate-1/2 CC \mathcal{B} can be generated by a generator matrix $\mathbf{G}(D) = [g_1(D), g_2(D)]$ [21, Ch. 4.2], where $g_1(D)$ and $g_2(D)$ are nonzero generator polynomials over the binary field¹. We assume that $g_1(D)$ defines odd bits of codewords $b_1[k]$, and $g_2(D)$ defines even bits $b_2[k]$. Any generator matrix $\mathbf{G}(D)$ can be put in a systematic form $\mathbf{G}_{\text{sys}}(D) = [1, g_2(D)/g_1(D)]$. Thus, an all-one input will produce a codeword where every odd bit is one, i.e., \mathbf{b}'' , such that $b_1''[k] = 1, \forall k$. Analogously, any generator matrix $\mathbf{G}(D)$ can be put in the form $\mathbf{G}'_{\text{sys}}(D) = [g_1(D)/g_2(D), 1]$, which means that an all-one input produces a codeword where every even bit is one, i.e., \mathbf{b}''' , such that $b_2'''[k] = 1, \forall k$. The three generator matrices $\mathbf{G}(D)$, $\mathbf{G}_{\text{sys}}(D)$, and $\mathbf{G}'_{\text{sys}}(D)$ generate the same code, i.e., any rate-1/2 CC \mathcal{B} satisfies the conditions of Corollary 2. This completes the proof. \square

Remark 2: Using a similar argument to the proof of Theorem 2, we can show that for codes satisfying conditions in Corollary 2, $w_c(\mathbf{x}, \hat{\mathbf{x}}) = 0$ not only for codewords at minimum distance but also for the first eight terms in the distance spectrum. We therefore conclude that the bound developed in [8] is, in fact, a TCM union bound (at least for the first 8 terms) obtained from the spectrum of a *binary* code.

C. Application: Optimal Bit-Wise Schemes

In this section, we show how optimal bit-wise schemes can be found for rate-1/2 CCs. One approach is presented in [8], where a search over all feedforward encoders was performed. The alternative approach we use here is to exploit the encoder equivalence shown in [9], which states that for CCs, different labelings can be grouped into classes that result in the same CM code \mathcal{X} . In other words, the same CM code \mathcal{X} can be obtained by any labeling within a class used together with a properly modified convolutional encoder. This allows us to use the results reported in [9] with the set-partitioning (SP) labeling [2].

For many constellations, including 4-PAM, the SP and Gray labelings belong to the same class [9, Theorem 3]. Let \mathcal{X}

¹We assume that any CC is realizable (see [21, Ch. 4.2]) and such that $g_i(D) \neq 0$ for $i = 1, 2$.

TABLE IV
GENERATOR POLYNOMIALS FOR RATE-1/2 CCs THAT GIVE OPTIMAL TCM ENCODERS FOR 4-PAM WITH THE BRGC

ν	\mathbf{G}	ν	\mathbf{G}
1	[3, 2]	5	[55, 51]
2	[7, 5]	6	[107, 135]
3	[13, 17]	7	[313, 235]
4	[23, 33]	8	[677, 515]

be a CM code obtained by the CC with generator matrix $\mathbf{G}_{\text{SP}}(D) = [g_1(D), g_2(D)]$ and 4-PAM with the SP labeling given by $\mathbf{q}_{\text{SP}} = [0, 1, 2, 3]$. The same CM code \mathcal{X} can be obtained by $\mathbf{G}_{\text{BRGC}}(D) = [g_1(D), g_1(D) + g_2(D)]$ and 4-PAM with GL_1 . We use this to obtain codes for the optimal bit-wise schemes, shown in Table IV, from codes for the optimal TCM schemes presented in [9, Table III]. From now on, we use octal representation for the generator polynomials and omit the argument D of the generator matrix. For memories $\nu = 2, 3, 4, 6, 7$, the codes in Table IV coincide with the codes in [8, Table III] ($\nu = 1, 8$ are not reported). For some ν , there may be several encoders with identical performance, which explains the different codes for $\nu = 5$.

Fig. 6 shows the S-DEC and the B-DEC performance for CCs with memories $\nu = 2, 4, 6, 8$ in Table IV. As predicted by the results in Sec. III-D, the B-DEC gives rise to a higher probability of error at moderate SNRs (the loss is approximately 0.2 dB). The gap between the B-DEC and the S-DEC decreases when the SNR increases, which is clearly seen from the curves marked with circles. As Fig. 5 suggests, the gap between the decoders is expected to be negligible at $d/\sigma_Z \approx 15$ dB. This corresponds to $E_s/N_0 \approx 11$ dB, which is beyond our simulation capabilities. To support the fact that the gap does indeed disappear at high SNR, in Fig. 7 we show ratios between the BER curves. As we can see, the curves behave similarly to the curve in Fig. 5, which confirms the asymptotic equivalence of the two decoders. At high SNR, the ratios in Fig. 7 can be interpreted as the loss in the pre-factor of the Q-function corresponding to the PEP for the codewords at minimum distance. Numerical analysis for different pairs of codewords, similar to the one in Sec. III-C, shows that at asymptotically high SNR, the ratios in Fig. 7 in fact converge to one, similarly to the ratio in Fig. 5, i.e., the loss in the pre-factor disappears.

V. CONCLUSIONS

In this paper, we compared the ML symbol-wise decoder and a suboptimal bit-wise decoder based on max-log L-values. It was shown that asymptotically, the loss caused by the use of the suboptimal bit-wise decoder is bounded, and in many cases equal to zero. The bit-wise decoder studied in this paper corresponds to the bit-interleaved coded modulation paradigm and is widely used in many wireless communication standards. The results in this paper can be seen as a theoretical justification for its use.

The analysis presented in this paper considered a 16-QAM constellation labeled with any Gray labeling. By means of numerical simulations, we have also studied two-dimensional

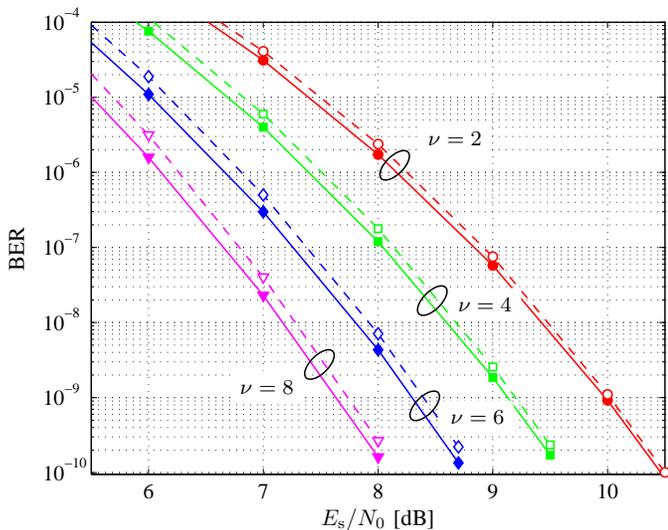


Fig. 6. BER simulation results for rate-1/2 CCs in Table IV over the AWGN channel. the S-DEC and the B-DEC are shown with solid and dashed lines, respectively.

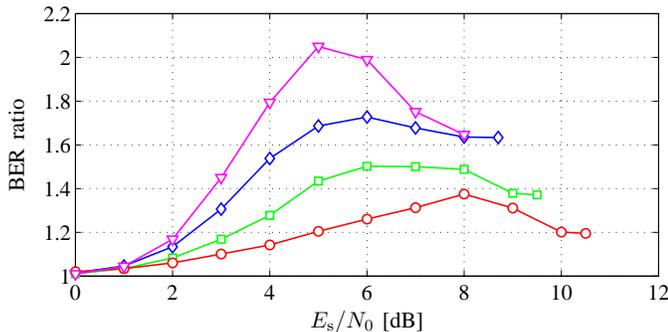


Fig. 7. Ratios between the BER curves for the B-DEC and the S-DEC in Fig. 6.

modulations such as 8-PSK and 64-QAM (not included in the paper). The results in this case also suggest that the asymptotic loss between symbol-wise and bit-wise decoders may be bounded. A rigorous analysis of these and other multilevel modulations is left for future investigation.

REFERENCES

- [1] G. Ungerboeck and I. Csajka, "On improving data-link performance by increasing channel alphabet and introducing sequence decoding," in *International Symposium on Information Theory (ISIT)*, Ronneby, Sweden, June 1976, (Book of abstracts).
- [2] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 1, pp. 55–67, Jan. 1982.
- [3] H. Imai and S. Hiraoka, "A new multilevel coding method using error-correcting codes," vol. IT-23, no. 3, pp. 371–377, May 1977.
- [4] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 927–946, May 1992.
- [5] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.
- [6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
- [7] C. Stierstorfer, R. F. H. Fischer, and J. B. Huber, "Optimizing BICM with convolutional codes for transmission over the AWGN channel," in *International Zurich Seminar on Communications*, Zurich, Switzerland, Mar. 2010.

- [8] A. Alvarado, L. Szczecinski, and E. Agrell, "On BICM receivers for TCM transmission," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2692–2707, Oct. 2011.
- [9] A. Alvarado, A. Graell i Amat, F. Brännström, and E. Agrell, "On optimal TCM encoders," *IEEE Trans. Commun.*, 2013 (to appear), available at <http://arxiv.org/abs/1210.2107>.
- [10] IEEE 802.11, "Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE Std 802.11-2012, Tech. Rep., Mar. 2012.
- [11] ETSI, "LTE; Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation," ETSI, Tech. Rep. ETSI TS 136 211 V11.2.0 (2013-04), Apr. 2013.
- [12] ETSI, "Digital video broadcasting (DVB); Frame structure channel coding and modulation for a second generation digital terrestrial television broadcasting system (DVB-T2)," ETSI, Tech. Rep. ETSI EN 302 755 V1.3.1 (2012-04), Apr. 2012.
- [13] R. D. Wesel, X. Liu, J. M. Cioffi, and C. Kominakis, "Constellation labeling for linear encoders," vol. 47, no. 6, pp. 2417–2431, Sep. 2001.
- [14] F. Gray, "Pulse code communications," U. S. Patent 2632058, Mar. 1953.
- [15] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.
- [16] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "Gray coding for multilevel constellations in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.
- [17] 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, pp. 2756–2765, June 2009.
- [18] M. Benjillali, L. Szczecinski, S. Aissa, and C. Gonzalez, "Evaluation of bit error rate for packet combining with constellation rearrangement," *Wiley Journal Wireless Comm. and Mob. Comput.*, vol. 8, no. 7, pp. 831–844, Sep. 2008.
- [19] A. Alvarado, L. Szczecinski, R. Feick, and L. Ahumada, "Distribution of L-values in Gray-mapped M^2 -QAM: Closed-form approximations and applications," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 2071–2079, July 2009.
- [20] E. Biglieri, D. Divsalar, P. J. McLane, and M. K. Simon, *Introduction to Trellis-Coded Modulation with Applications*. Macmillan, 1991.
- [21] W. E. Ryan and S. Lin, *Channel codes: Classical and Modern*, 1st ed. Cambridge University Press, 2009.

Mikhail Ivanov was born in St. Petersburg, Russia. He received his M.Sc. in Engineering and Technologies from St. Petersburg State Electrotechnical University "LETI", Russia, in 2009, and the degree of Licentiate of Engineering (Teknologie Licentiatexamen) from Chalmers University of Technology, Gothenburg, Sweden, in 2013. He is currently working towards the Ph.D. degree at the Dept. of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. His research interests include channel coding, modulation, and information theory, as well as coded random access techniques in vehicular ad-hoc networks.

Alex Alvarado (S'06–M'11) was born in Quellón, on the island of Chiloé, Chile. He received his Electronics Engineer degree (Ingeniero Civil Electrónico) and his M.Sc. degree (Magíster en Ciencias de la Ingeniería Electrónica) from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2003 and 2005, respectively. He obtained the degree of Licentiate of Engineering (Teknologie Licentiatexamen) in 2008 and his PhD degree in 2011, both of them from Chalmers University of Technology, Gothenburg, Sweden.

From 2011 to 2012, he was a Newton International Fellow at the University of Cambridge, United Kingdom, funded by The British Academy and The Royal Society. He is currently a Marie Curie Intra-European Fellow at the same institution. His general research interests are in the areas of digital communications, coding, and information theory, and in particular, in the design of coded modulation systems.

Dr. Alvarado was a holder of the Merit Scholarship Program for Foreign Students, granted by the Ministère de l'Éducation, du Loisir et du Sports du Québec. He is a recipient of the IEEE 2013 Communication Theory Workshop (CTW) Best Poster Award.

Fredrik Brännström (S'98–M'05) received the M.Sc. degree in Electrical Engineering from Luleå University of Technology (Luleå, Sweden) in 1998, the Lic.Eng. degree in Communication Theory, and the Ph.D. degree in Communication Theory from the Department of Computer Engineering, Chalmers University of Technology (Gothenburg, Sweden) in 2000 and 2004, respectively. In 2012 he received the Docent title in Communication Systems from the Department of Signals and Systems, Chalmers University of Technology.

He has spent periods of 2001, 2002, 2003, and 2005 as a visiting researcher at the Institute for Telecommunications Research, University of South Australia (Adelaide, Australia). From 2004 to 2006, he was a Postdoctoral Researcher in the Communication Systems Group at the Department of Signals and Systems, Chalmers University of Technology. From 2006 to 2010 he was a Senior Algorithm Engineer and Principal Design Engineer at Quantenna Communications (Fremont, CA). At Quantenna he was part of the system team that developed the PHY layer for the worlds first WiFi 802.11n chip with 4x4 MIMO using dynamic digital beamforming together with LDPC codes operating at 600 Mbps. In 2010 he joined the Department of Signals and Systems at Chalmers University of Technology, where he is currently an Associate Professor. His research interests in communication theory and information theory include code design, coded modulation, bit-to-symbol mappings, and efficient iterative processing, as well as MIMO signal processing and algorithm design for vehicular communication systems.

Erik Agrell (M'99–SM'02) received the Ph.D. degree in information theory in 1997 from Chalmers University of Technology, Sweden.

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, where he is a Professor in Communication Systems since 2009. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the signals and systems research area. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications.

Prof. Agrell served as Publications Editor for the IEEE Transactions on Information Theory from 1999 to 2002 and is an Associate Editor for the IEEE Transactions on Communications since 2012. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, the 2013 CTW Best Poster Award, and the 2013 Chalmers Supervisor of the Year Award.