High-SNR Asymptotics of Mutual Information for Discrete Constellations

High-SNR Asymptotics of Mutual Information for Discrete Constellations with Applications to BICM

Alex Alvarado, Fredrik Brännström, Erik Agrell, and Tobias Koch

***Abstract*—The high-signal-to-noise ratio (SNR) asymptotic behavior of the mutual information (MI) for discrete constellations over the scalar additive white Gaussian noise channel is studied. Exact asymptotic expressions for the MI for arbitrary one-dimensional constellations and input distributions are presented in the limit as the SNR tends to infinity. Asymptotics of the MMSE and symbol-error probability (SEP) are also developed. It is shown that for any input distribution, the MI, MMSE and SEP have an asymptotic behavior proportional to the Gaussian Q-function, whose argument depends only on the minimum Euclidean distance of the constellation and the SNR, and where the proportionality constants are functions of the number of pairs of constellation points at minimum Euclidean distance and their corresponding probabilities. Closed-form expressions for the coefficients of these Q-functions are presented. The developed expressions are used to study the high-SNR behavior of the generalized mutual information (GMI) for bit-interleaved coded modulation (BICM). In particular, the long-standing conjecture that Gray codes are the binary labelings that maximize the BICM-GMI at high SNR is proven. It is also shown that for any equally spaced constellation whose size is a power of two, there always exists an anti-Gray code that gives the lowest BICM-GMI at high SNR.**

***Index Terms*—Anti-Gray code, additive white Gaussian noise channel, bit-interleaved coded modulation, discrete constellations, Gray code, minimum-mean square error, mutual information, high-SNR asymptotics.**

I. Introduction

In this paper we consider the real additive white Gaussian noise (AWGN) channel

$$Y = \sqrt{\rho}X + Z \tag{1}$$

where X is the transmitted symbol and Z is a Gaussian random variable, independent of X, with zero mean and unit variance. The capacity of the real AWGN channel in (1) is given by [1]

$$\mathsf{C}(\rho) = \frac{1}{2}\log(1+\gamma) \tag{2}$$

Research supported by the European Community's Seventh's Framework Programme (FP7/2007-2013) under grant agreement No. 271986, by the Swedish Research Council, Sweden (under grants #621-2006-4872 and #621-2011-5950) and by Ministerio de Economía of Spain (projects "DEIPRO", id. TEC2009-14504-C02-01, and "COMONSENS", id. CSD2008-00010). Parts of this work will be presented at the IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013.

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

E. Agrell and F. Brännström are with the Dept. of Signals and Systems, Chalmers Univ. of Technology, SE-41296 Göteborg, Sweden (email: {fredrik.brannstrom, agrell}@chalmers.se).

T. Koch is with the Department of Signal Theory and Communications, Universidad Carlos III de Madrid, 28911 Leganés, Spain (email: koch@tsc.uc3m.es).

where $\gamma \triangleq \rho\mathbb{E}_X[X^2]$ is the signal-to-noise ratio (SNR) and $\rho > 0$ is an arbitrary scale factor. Although inputs distributed according to the Gaussian distribution attain the capacity, they suffer from several drawbacks which prevent them from being used in practical systems. Among them, especially relevant are the unbounded support and the infinite number of bits needed to represent signal points. In practical systems, discrete distributions are typically preferred.

The mutual information (MI) between the channel input X and the channel output Y of (1), where the input distribution is constrained to be a probability mass function (PMF) over a discrete constellation, represents the maximum rate at which information can be reliably transmitted over (1) using that particular constellation. While the low-SNR asymptotics of the MI for discrete constellations are well understood (see [1]–[4] and references therein), to the best of our knowledge, only upper and lower bounds are known for the high-SNR behavior [5]–[7]. It was observed in [6, p. 1073] that for discrete constellations, maximizing the MI is equivalent to minimizing either the symbol-error probability (SEP) or the minimum mean-square error (MMSE). In [8, Appendix E], two constellations with different minimum Euclidean distances (MEDs) are compared, and it is shown that, for sufficiently large SNR, the constellation with larger MED gives a higher MI. Upper and lower bounds on the MI and MMSE for multiple-antenna systems over fading channels can be found in [9]–[11]. Using the Mellin transform method, asymptotic expansions for the MMSE and MI for scalar and vectorial coherent fading channels were recently derived in [12].

In this paper, we study high-SNR asymptotics of the MI for discrete constellations. In particular, we consider arbitrary constellations and input distributions (independent of ρ) and find exact asymptotic expressions for the MI in the limit as the SNR tends to infinity. Exact asymptotic expressions for the MMSE and SEP are also developed. We prove that for any constellation and input distribution, the MI, MMSE, and SEP have an asymptotic behavior proportional to $\mathsf{Q}\left(\sqrt{\rho}d/2\right)$, where $\mathsf{Q}(\cdot)$ is the Gaussian Q-function and d is the MED of the constellation. While this asymptotic behavior has been demonstrated for uniform input distributions (e.g., [6, eqs. (36)–(37)], [6, Sec. II-C], [9, Sec. III], [11, Sec. III]), we show that it holds for any discrete input distribution that does not depend on the SNR. Furthermore, in contrast to previous works, we provide closed-form expressions for the coefficients before the Q-functions, thereby characterizing the asymptotic behavior of the MI, MMSE, and SEP more accurately.

While these asymptotical results are general, we use them to study bit-interleaved coded modulation (BICM) [13]–[15],

which can be viewed as a pragmatic approach for coded modulation [15, Ch. 1]. The key element in BICM is the use of a (suboptimal) bit-wise detection rule, which was cast as a mismatched decoder in [16]. BICM is used in many of the current wireless communications standards, e.g., HSPA, IEEE 802.11a/g/n, and the DVB standards (DVB-T2/S2/C2).

The BICM generalized mutual information (BICM-GMI) is an achievable rate for BICM [16] and depends heavily on the binary labeling of the constellation. The optimality of a Gray code (GC) in terms of maximizing the BICM-GMI was conjectured in [14, Sec. III-C]; however, it was shown in [17] that for low and medium SNRs, there exist other labelings that give a higher BICM-GMI (see also [18, Ch. 3]). For further results on BICM at low SNR see [19]–[22]. On the other hand, numerical results presented in [18, Ch. 3] and [23] suggest that GCs are optimal at high SNR in terms of BICM-GMI. However, to the best of our knowledge, the optimality of GCs at high SNR has never been proven.

In this paper, we derive an asymptotic expression for the BICM-GMI as a function of the constellation, input distribution, and binary labeling. Using this expression, we then prove the optimality of GCs at high SNR. Using the MI-MMSE relationship, an asymptotic expression for the derivative of the BICM-GMI is also developed. The obtained asymptotic expressions for the BICM-GMI and its derivative, as well as the one for the bit-error probability (BEP), are all shown to be proportional to $\mathsf{Q}\left(\sqrt{\rho}d/2\right)$.

This paper is organized as follows. In Sec. II, the notation convention and system model are presented. The asymptotics of the MI and MMSE are presented in Sec. III and BICM is studied in Sec. IV. The conclusions are drawn in Sec. V.

II. Preliminaries

A. Notation Convention

Row vectors are denoted by boldface letters $\boldsymbol{x} = [x_1, x_2, \ldots, x_M]$ and sets are denoted by calligraphic letters $\mathcal{C}$. An exception is the set of real numbers, which is denoted by $\mathbb{R}$. The binary set is defined as $\mathcal{B} \triangleq \{0, 1\}$ and the bipolar set as $\mathcal{W} \triangleq \{-1, +1\}$. The negation of a bit b is denoted by $\overline{b}$. All the logarithms are natural logarithms and all the MIs are therefore given in nats. Probability density functions (PDFs) and conditional PDFs are denoted by $f_Y(y)$ and $f_{Y|X}(y|x)$, respectively. Analogously, PMFs are denoted by $P_X(x)$ and $P_{X|Y}(x|y)$. Expectations over a random variable X are denoted by $\mathbb{E}_X[\cdot]$.

B. Model

We consider the discrete-time, real-valued AWGN channel in (1), where the transmitted symbols X are constrained to $X \in \mathcal{X} \triangleq \{x_1, x_2, \ldots, x_M\}$ and $|\mathcal{X}| = M = 2^m$. The set of indices that enumerates all the constellation symbols in $\mathcal{X}$ is defined as $\mathcal{I}_\mathcal{X} \triangleq \{1, \ldots, M\}$.

We focus on one-dimensional constellations and assume, without loss of generality, that the symbols are different and ordered, i.e., $x_1 < x_2 < \cdots < x_M$. Each of the symbols is transmitted with probability $p_i \triangleq P_X(x_i)$, $0 < p_i < 1$. While the transmitted symbols are fully determined by the PMF P_X, we shall use *constellation* to denote the support $\mathcal{X}$ of the PMF and *input distribution* to denote the probabilities $\boldsymbol{p} = [p_1, \ldots, p_M]$ associated with the symbols. We assume that neither the constellation nor the input distribution depends on ρ.

The transmitted average symbol energy is finite and given by

$$E_\mathrm{s} \triangleq \mathbb{E}_X[X^2] = \sum_{i \in \mathcal{I}_\mathcal{X}} p_i x_i^2. \tag{3}$$

It follows that the SNR γ in (2) is $\gamma = \rho E_\mathrm{s}$.

An M-ary pulse-amplitude modulation (MPAM) constellation having M equally spaced symbols (separated by 2Δ) is denoted by $\mathcal{E} \triangleq \{x_i = -(M - 2i + 1)\Delta : i = 1, \ldots, M\}$. A uniform distribution of X is denoted by P_X^u, i.e., $p_i = 1/M$ $\forall i$. A uniform input distribution with $\mathcal{X} = \mathcal{E}$ is denoted by P_X^eu, where in this case $\Delta^2 = 3E_\mathrm{s}/(M^2 - 1)$.

The Gaussian Q-function is defined as

$$\mathsf{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-\frac{1}{2}\xi^2} \, \mathrm{d}\xi \tag{4}$$

the entropy of the random variable X as

$$H_{P_X} \triangleq -\mathbb{E}_X\left[\log\left(P_X(X)\right)\right] \tag{5}$$

the MI between X and Y as

$$\mathsf{I}_{P_X}(\rho) \triangleq \mathbb{E}_{X,Y}\left[\log\left(f_{Y|X}(Y|X)/f_Y(Y)\right)\right] \tag{6}$$

and the MMSE as

$$\mathsf{M}_{P_X}(\rho) \triangleq \mathbb{E}_{X,Y}[(X - \hat{X}^\mathrm{ME}(Y))^2] \tag{7}$$

where $\hat{X}^\mathrm{ME}(y) \triangleq \mathbb{E}_X[X|Y = y]$ is the conditional (posterior) mean estimator.

We also define the SEP as

$$\mathsf{S}_{P_X}(\rho) \triangleq \Pr\{\hat{X}^\mathrm{MAP}(Y) \neq X\} \tag{8}$$

where X is the transmitted symbol and

$$\hat{X}^\mathrm{MAP}(y) \triangleq \operatorname*{argmax}_{x \in \mathcal{X}} P_{X|Y}(x|y) \tag{9}$$

is the decision made by a maximum a posteriori probability (MAP) symbol demapper.

C. Discrete Constellations

The MED of the constellation is defined as

$$d \triangleq \min_{x_i, x_j \in \mathcal{X} : i \neq j} |x_i - x_j|. \tag{10}$$

We define the counting function

$$A_\mathcal{X}^{(i)}(\delta) \triangleq \begin{cases} 1, & \text{if } \exists x \in \mathcal{X} : x_i - x = \delta \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where $\delta \in \mathbb{R}$. Since $x_i \in \mathcal{X}$, we have $A_\mathcal{X}^{(i)}(0) = 1$ $\forall i \in \mathcal{I}_\mathcal{X}$. We further define $A_\mathcal{X}$ as twice the number of pairs of constellation points at MED, i.e.,

$$A_\mathcal{X} \triangleq \sum_{i \in \mathcal{I}_\mathcal{X}} \sum_{w \in \mathcal{W}} A_\mathcal{X}^{(i)}(wd). \tag{12}$$

By using the fact that for any real-valued constellation there are at least one and at most $M-1$ pairs of constellation points at MED, we obtain the bound

$$2 \leq A_{\mathcal{X}} \leq 2(M-1). \tag{13}$$

The upper bound is achieved by an MPAM constellation, for which

$$A_{\mathcal{E}} = 2(M-1). \tag{14}$$

Analogous to $A_{\mathcal{X}}^{(i)}(\delta)$, we define $B_{P_X}^{(i)}(\delta)$ as

$$B_{P_X}^{(i)}(\delta) \triangleq \begin{cases} \sqrt{p_j p_i}, & \text{if } \exists x_j \in \mathcal{X} : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases}. \tag{15}$$

Clearly $B_{P_X}^{(i)}(0) = p_i, \forall i \in \mathcal{I}_{\mathcal{X}}$.

Finally, for a given P_X, we define the constant

$$B_{P_X} \triangleq \sum_{i \in \mathcal{I}_{\mathcal{X}}} \sum_{w \in \mathcal{W}} B_{P_X}^{(i)}(wd). \tag{16}$$

For a uniform input distribution, $P_X = P_X^{\mathrm{u}}$ and $M B_{P_X}^{(i)}(\delta) = A_{\mathcal{X}}^{(i)}(\delta)$, so

$$B_{P_X^{\mathrm{u}}} = \frac{A_{\mathcal{X}}}{M}. \tag{17}$$

Example 1: Consider an unequally spaced 4-ary constellation with $x_1 = -4$, $x_2 = -2$, $x_3 = 2$, and $x_4 = 4$, and the input distribution $p_i = i/10$ with $i = 1, 2, 3, 4$. The MED in (10) is $d = 2$, E_{s} in (3) is $E_{\mathrm{s}} = 10$, $A_{\mathcal{X}}$ in (12) is $A_{\mathcal{X}} = 4$ (two pairs of constellation points at MED), and B_{P_X} in (16) is $B_{P_X} = 2\sqrt{p_1 p_2} + 2\sqrt{p_3 p_4} \approx 0.98$. This example will be continued in Example 4.

III. High-SNR Asymptotics

There exists a fundamental relationship between the MI and the MMSE for AWGN channels [24] (see also [25, Ch. 2]):

$$\frac{\mathrm{d}}{\mathrm{d}\rho} \mathsf{I}_{P_X}(\rho) = \frac{1}{2} \mathsf{M}_{P_X}(\rho). \tag{18}$$

Exploiting this MI-MMSE relation, bounds on the MI can be used to derive bounds on the MMSE and *vice versa*.

Upper and lower bounds on the MI and MMSE for discrete constellations at high SNR can be found, e.g., in [5]–[7], [9]–[12]. While these bounds describe the correct asymptotic behavior, they are, in general, not tight in the sense that the ratio between them does not tend to one as $\rho \to \infty$. In what follows, we present exact asymptotic expressions for the MI and MMSE for any arbitrary P_X.

A. Asymptotics of the MI, MMSE, and SEP

For any given input distribution P_X, the MI tends to H_{P_X} as ρ tends to infinity. In the following we study how fast the MI converges towards its maximum H_{P_X} by analyzing the difference $H_{P_X} - \mathsf{I}_{P_X}(\rho)$.[1] Theorem 1 is the main result of this paper and characterizes the high-SNR behavior of $H_{P_X} - \mathsf{I}_{P_X}(\rho)$.

Theorem 1: For any P_X

$$\lim_{\rho \to \infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho} d/2\right)} = \pi B_{P_X} \tag{19}$$

where B_{P_X} is given by (16).

Proof: The proof is given in Appendix A. □

Similar to Theorem 1, we have the following asymptotic expression for the MMSE.

Theorem 2: For any P_X

$$\lim_{\rho \to \infty} \frac{\mathsf{M}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho} d/2\right)} = \frac{\pi d^2}{4} B_{P_X} \tag{20}$$

where B_{P_X} is given by (16).

Proof: The proof is given in Appendix B. □

In analogy to Theorems 1 and 2, an asymptotic expression for the SEP can be obtained.

Theorem 3: For any P_X

$$\lim_{\rho \to \infty} \frac{\mathsf{S}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho} d/2\right)} = B_{P_X} \tag{21}$$

where B_{P_X} is given by (16).

Proof: The proof is given in Appendix C. □

Theorems 1–3 reveal that, at high SNR, the MI, MMSE, and SEP behave as

$$\mathsf{I}_{P_X}(\rho) \approx H_{P_X} - \pi B_{P_X} \mathsf{Q}\left(\frac{\sqrt{\rho} d}{2}\right), \tag{22}$$

$$\mathsf{M}_{P_X}(\rho) \approx \frac{\pi d^2}{4} B_{P_X} \mathsf{Q}\left(\frac{\sqrt{\rho} d}{2}\right), \tag{23}$$

$$\mathsf{S}_{P_X}(\rho) \approx B_{P_X} \mathsf{Q}\left(\frac{\sqrt{\rho} d}{2}\right). \tag{24}$$

The results in (22)–(24) show that for any input distribution, the MI, MMSE, and SEP have the same high-SNR behavior, i.e., they are all proportional to a Gaussian Q-function,[2] where the proportionality constants depend on the input distribution and, in the case of the MMSE, also on the MED of the constellation. Hence, the one-dimensional constellation that maximizes the MI is the same one that minimizes both the SEP and the MMSE.

Remark 1: While the results presented in this section hold for one-dimensional constellations, they directly generalize to multidimensional constellations that are constructed as *ordered direct products* [21, eq. (1)] of one-dimensional constellations. For example, the results directly generalize to rectangular quadrature amplitude modulation constellations.

B. Discussion and Examples

For a uniform input distribution ($P_X = P_X^{\mathrm{u}}$), Theorems 1–3 particularize to the following result.

[1] The quantity $H_{P_X} - \mathsf{I}_{P_X}(\rho)$ is the conditional entropy of X given Y.

[2] Disregarding the "offset" H_{P_X} in (22).

TABLE I
SUMMARY OF ASYMPTOTICS OF MI, MMSE, AND SEP.

Input Distribution	P_X	P_X^{u}	P_X^{eu}
$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	πB_{P_X}	$\pi \frac{A_{\mathcal{X}}}{M}$	$\frac{2\pi(M-1)}{M}$
$\lim_{\rho\to\infty} \frac{\mathsf{M}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	$\frac{\pi d^2}{4} B_{P_X}$	$\frac{\pi d^2}{4}\frac{A_{\mathcal{X}}}{M}$	$\frac{6\pi E_{\mathrm{s}}}{M(M+1)}$
$\lim_{\rho\to\infty} \frac{\mathsf{S}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	B_{P_X}	$\frac{A_{\mathcal{X}}}{M}$	$\frac{2(M-1)}{M}$

Corollary 1: For any $\mathcal{X}$ with a uniform input distribution

$$\lim_{\rho\to\infty} \frac{\log M - \mathsf{I}_{P_X^{\mathrm{u}}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \pi \frac{A_{\mathcal{X}}}{M}, \tag{25}$$

$$\lim_{\rho\to\infty} \frac{\mathsf{M}_{P_X^{\mathrm{u}}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4}\frac{A_{\mathcal{X}}}{M}, \tag{26}$$

$$\lim_{\rho\to\infty} \frac{\mathsf{S}_{P_X^{\mathrm{u}}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{A_{\mathcal{X}}}{M} \tag{27}$$

where $A_{\mathcal{X}}$ is given in (12).

Proof: From Theorems 1–3 and (17). □

The expression (27) corresponds to the well-known high-SNR approximation for the SEP [26, eq. (2.3-29)]. Moreover, Corollary 1 shows that for a uniform input distribution, the MI, the MMSE, and the SEP for discrete constellations in the high-SNR regime are functions of the MED of the constellation and the number of pairs of constellation points at MED only.

For MPAM and a uniform input distribution ($P_X = P_X^{\mathrm{eu}}$), Corollary 1 particularizes to (see (14))

$$\mathsf{I}_{P_X^{\mathrm{eu}}}(\rho) \approx \log M - \frac{2\pi(M-1)}{M}\mathsf{Q}\left(\sqrt{\rho}d/2\right), \tag{28}$$

$$\mathsf{M}_{P_X^{\mathrm{eu}}}(\rho) \approx \frac{6\pi E_{\mathrm{s}}}{M(M+1)}\mathsf{Q}\left(\sqrt{\rho}d/2\right), \tag{29}$$

$$\mathsf{S}_{P_X^{\mathrm{eu}}}(\rho) \approx \frac{2(M-1)}{M}\mathsf{Q}\left(\sqrt{\rho}d/2\right). \tag{30}$$

In Table I, the results obtained in Theorems 1–3, Corollary 1, and (28)–(30) are summarized.

Example 2: In Fig. 1, we show the conditional entropy $\log M - \mathsf{I}_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM with uniform input distributions[3] together with the asymptotic expression in (28). We also show the lower and upper bounds derived in [6, eq. (34)–(35)] and [11, eq. (17)–(19)]. Observe that (28) approximates $\mathsf{I}_{P_X^{\mathrm{eu}}}(\rho)$ accurately for a large range of SNR. In Fig. 2, analogous results for the MMSE are presented, where the bounds derived in [6, eq. (30)–(31)] and [11, eq. (13)–(15)] are also included. Also here our asymptotic expression (29) approximates the MMSE accurately for a large range of SNR.

Remark 2: It follows from Corollary 1 that the constellation that maximizes the MI (or equivalently, the constellation that minimizes the MMSE and the SEP) at high SNR is the constellation that first maximizes the MED and then minimizes $A_{\mathcal{X}}$. For one-dimensional constellations with the same E_{s}, the MED is maximized by an MPAM constellation ($\mathcal{X} = \mathcal{E}$).

We conclude this section by noting that if Theorems 1 and 2 are combined, we obtain

$$\lim_{\rho\to\infty} \frac{\mathsf{M}_{P_X}(\rho)}{H_{P_X} - \mathsf{I}_{P_X}(\rho)} = \frac{d^2}{4}. \tag{31}$$

Thus, for any P_X, the limiting ratio between the MMSE and the conditional entropy does not depend on the input distribution. Moreover, using Theorems 1 and 3, we obtain

$$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{\mathsf{S}_{P_X}(\rho)} = \pi. \tag{32}$$

[3]Calculated numerically using Gauss–Hermite quadratures with 300 quadrature points [23, Sec. III].

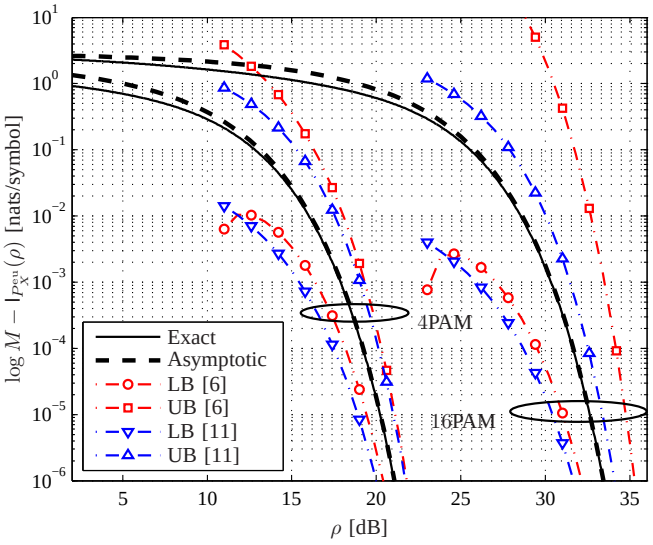

Fig. 1. $\log M - \mathsf{I}_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_{\mathrm{s}} = 1$) and the asymptotic expression in (28) (thick dashed lines). The lower and upper bounds [6, eq. (34)–(35)] and [11, eq. (17)–(19)] are also shown.

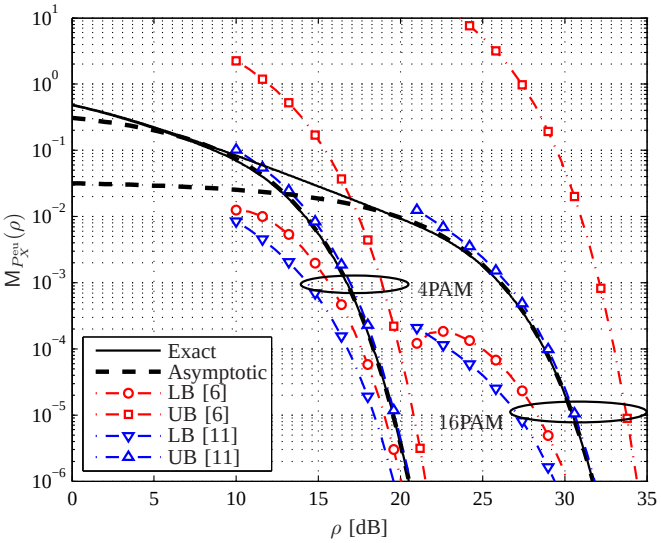

Fig. 2. $\mathsf{M}_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_{\mathrm{s}} = 1$) and the asymptotic expression in (29) (thick dashed lines). The lower and upper bounds [6, eq. (30)–(31)] and [11, eq. (13)–(15)] are also shown.

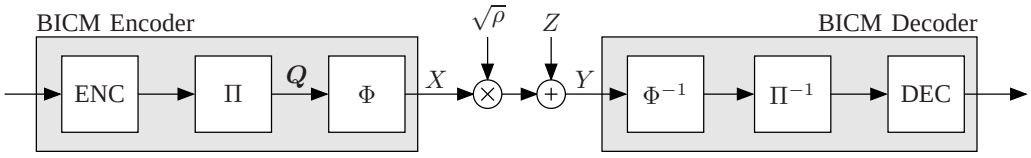

Fig. 3. A BICM scheme: The BICM encoder is formed by a serial concatenation of a binary encoder (ENC), a bit-level interleaver (Π), and a memoryless mapper (Φ). The BICM decoder is based on a demapper (Φ^{-1}) that computes logarithmic likelihood ratios, a de-interleaver (Π^{-1}), and a channel decoder (DEC).

Thus, for any P_X, the limiting ratio between the conditional entropy and the SEP equals π.

IV. Application: Binary Labelings for Bit-Interleaved Coded Modulation

In BICM [13]–[15] (see Fig. 3), the encoder is realized as a serial concatenation of a binary encoder, a bit-level interleaver, and a memoryless mapper. At the receiver's side, the demapper computes logarithmic likelihood ratios, which are de-interleaved and then decoded. A key element for the performance of BICM is $\Phi : \mathcal{B}^m \to \mathcal{X}$, which maps coded bits to constellation symbols. In this section we study the high-SNR behavior of BICM. Using the results in Sec. III, we will find an asymptotic expression for the BICM-GMI and we will study the relationship between the BICM-GMI and the BEP. We will also prove that GCs are optimal in terms of BICM-GMI for one-dimensional constellations with uniform input distributions.

A. BICM Model

A binary labeling for a constellation is defined by the vector $\boldsymbol{l} = [l_1, l_2, \ldots, l_M]$ where $l_i \in \{0, 1, \ldots, M-1\}$ is the integer representation of the ith length-m binary label $\boldsymbol{q}_i = [q_{i,1}, \ldots, q_{i,m}] \in \mathcal{B}^m$ associated with the symbol x_i, with $q_{i,1}$ being the most significant bit. The labeling defines $2m$ *subconstellations* $\mathcal{X}_{k,b} \subset \mathcal{X}$ for $k = 1, \ldots, m$ and $b \in \mathcal{B}$, given by $\mathcal{X}_{k,b} \triangleq \{x_i \in \mathcal{X} : q_{i,k} = b\}$ with $|\mathcal{X}_{k,b}| = M/2$. We define $\mathcal{I}_{\mathcal{X}_{k,b}} \subset \{1, \ldots, M\}$ as the indices of the symbols in $\mathcal{X}$ that belong to $\mathcal{X}_{k,b}$.

Example 3: In Fig. 4, we show the $2m = 6$ subconstellations for an 8PAM constellation labeled by the binary reflected Gray code (BRGC) $\boldsymbol{l} = [0, 1, 3, 2, 6, 7, 5, 4]$ [27]–[29], as well as the corresponding values of $\mathcal{I}_{\mathcal{X}_{k,b}}$ and $A_{\mathcal{X}_{k,b}}$.

In BICM, the coded bits $\boldsymbol{Q} = [Q_1, Q_2, \ldots, Q_m]$ at the input of the mapper (see Fig. 3) are assumed to be independent but possibly nonuniformly distributed. Therefore, the vector of bit probabilities $[P_{Q_1}(0), P_{Q_2}(0), \ldots, P_{Q_m}(0)]$ induces a symbol input distribution P_X via the labeling as [21, eq. (31)] [30, eq. (8)]

$$P_X(x_i) = p_i = \prod_{k=1}^{m} P_{Q_k}(q_{i,k}). \tag{33}$$

Using (33), we obtain the conditional probabilities

$$P_{X|Q_k}(x|b) = \begin{cases} \frac{P_X(x)}{P_{Q_k}(b)}, & \text{if } x \in \mathcal{X}_{k,b} \\ 0, & \text{if } x \notin \mathcal{X}_{k,b} \end{cases} \tag{34}$$

for $k = 1, \ldots, m$ and $b \in \mathcal{B}$. According to (34), each of the $2m$ conditional input distributions $[P_{X|Q_k}(x_1|b), \ldots, P_{X|Q_k}(x_M|b)]$ has $M/2$ non-zero probabilities, which specify which of the $M/2$ symbols in $\mathcal{X}$ are included in $\mathcal{X}_{k,b}$.

For uniformly distributed bits, i.e., $P_{Q_k}(b) = 1/2$, it follows that the symbol distribution is also uniform, i.e., $P_X = P_X^{\mathrm{u}}$, and thus,

$$P_{X|Q_k}(x_i|b) = \begin{cases} \frac{2}{M}, & \text{if } x_i \in \mathcal{X}_{k,b} \\ 0, & \text{if } x_i \notin \mathcal{X}_{k,b} \end{cases}. \tag{35}$$

We shall use $X_{k,b}$ to denote a random variable with support $\mathcal{X}_{k,b}$ and probabilities $P_{X|Q_k}(x|b)$ for $x \in \mathcal{X}_{k,b}$ in (34). The corresponding PMF is denoted by $P_{X_{k,b}}$ and the PMF for the uniform case in (35) is denoted by $P^{\mathrm{u}}_{X_{k,b}}$.

In what follows, we will apply the results of Sec. III to BICM. To this end, we will often replace $\mathcal{X}$ and P_X in Sec. III by $\mathcal{X}_{k,b}$ and $P_{X_{k,b}}$, respectively. Note, however, that d as defined in (10), still denotes the MED Euclidean distance (ED) of the constellation $\mathcal{X}$. We will not consider the MED for subconstellations. This implies that it is possible that no pairs of constellation points in $\mathcal{X}_{k,b}$ are at MED. Consequently, the bounds on $A_{\mathcal{X}_{k,b}}$ are

$$0 \le A_{\mathcal{X}_{k,b}} \le 2\,(M/2 - 1) \tag{36}$$

which differ from the corresponding bounds on $A_{\mathcal{X}}$ in (13).

B. Binary Labelings for BICM

The *natural binary code (NBC)* [21, Sec. II-B] is defined as the binary labeling $\boldsymbol{l}$ where $l_i = i - 1$, for $i = 1, 2, \ldots, M$. The NBC is an important labeling for BICM because it is the unique optimal labeling for BICM in the low-SNR regime for $\mathcal{X} = \mathcal{E}$ [21, Theorem 14], [22]. A labeling $\boldsymbol{l}$ is said to be a GC if for all i, j such that $|x_i - x_j| = d$, the binary labels $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ are at Hamming distance one. One of the most popular GCs is the BRGC [27]–[29], which we showed in Example 3 for $M = 8$.

To characterize binary labelings we define the constant

$$C_{\mathcal{X},\boldsymbol{l}} \triangleq \sum_{k=1}^{m} \sum_{b \in \mathcal{B}} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w \in \mathcal{W}} A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(wd). \tag{37}$$

For a given subconstellation $\mathcal{X}_{k,b}$, the two inner sums in (37) consider all the constellation points in the subconstellation $\mathcal{X}_{k,\overline{b}}$ at MED from $x_i \in \mathcal{X}_{k,b}$. Thus, the quantity $C_{\mathcal{X},\boldsymbol{l}}$ corresponds to twice the total number of *different* bits between the labels of constellation symbol pairs at MED. Using this

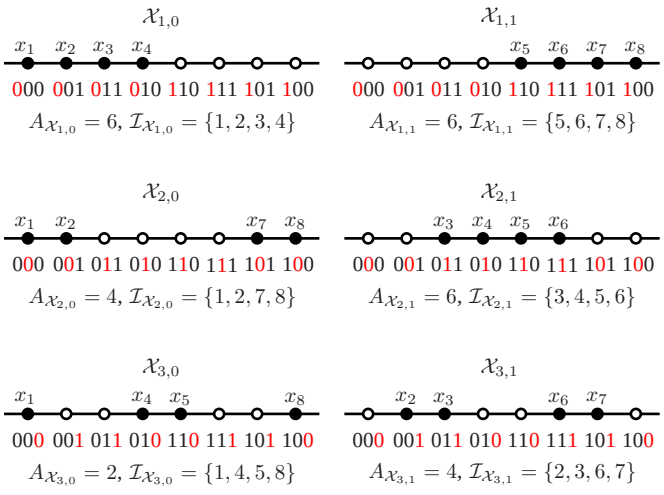

Fig. 4. Subconstellations $\mathcal{X}_{k,b}$ (black circles) for 8PAM labeled by the BRGC $\boldsymbol{l} = [0,1,3,2,6,7,5,4]$, where the values of $q_{i,k}$ for $k = 1,2,3$ are shown in red, and where $A_{\mathcal{X}} = C_{\mathcal{X},\boldsymbol{l}} = 14$. The values of $A_{\mathcal{X}_{k,b}}$ and $\mathcal{I}_{\mathcal{X}_{k,b}}$ are also shown.

interpretation, it follows that (37) can also be expressed as

$$C_{\mathcal{X},\boldsymbol{l}} = \sum_{k=1}^{m} \left(A_{\mathcal{X}} - A_{\mathcal{X}_{k,0}} - A_{\mathcal{X}_{k,1}} \right) \tag{38}$$

where $A_{\mathcal{X}} - A_{\mathcal{X}_{k,0}} - A_{\mathcal{X}_{k,1}}$ corresponds to twice the number of pairs of constellation points at MED with different labeling at bit position k. For example, for the constellation and labeling in Fig. 4, $C_{\mathcal{X},\boldsymbol{l}} = 14 = A_{\mathcal{X}}$.

While $A_{\mathcal{X}}$ in (12) depends only on the geometry of the constellation, $C_{\mathcal{X},\boldsymbol{l}}$ in (37) depends on both the geometry of the constellation and the labeling. By noting that any pair of constellation points at MED will differ in at least one bit, we obtain that for any $\mathcal{X}$ and $\boldsymbol{l}$

$$C_{\mathcal{X},\boldsymbol{l}} \geq A_{\mathcal{X}}. \tag{39}$$

For example, for $\mathcal{X} = \mathcal{E}$ and the NBC, $C_{\mathcal{X},\boldsymbol{l}}$ can be expressed as

$$\begin{aligned} C_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} &= 2\sum_{k=1}^{m} (2^k - 1) \\ &= 2(2M - m - 2) \end{aligned} \tag{40}$$

which is obtained by noting that, for each k, there are $2^k - 1$ symbols satisfying $q_{i,k} \neq q_{i+1,k}$, for $i = 1,2,\ldots,M-1$.

We also define the constant

$$D_{P_X,\boldsymbol{l}} \triangleq \sum_{k=1}^{m} \sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w\in\mathcal{W}} D^{(i)}_{P_{X_{k,\bar{b}}}}(wd) \tag{41}$$

where $D^{(i)}_{P_{X_{k,b}}}(wd)$ with $b \in \mathcal{B}$ is defined as

$$D^{(i)}_{P_{X_{k,b}}}(\delta) \triangleq \begin{cases} \sqrt{p_j p_i}, & \text{if } \exists x_j \in \mathcal{X}_{k,b} : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases}. \tag{42}$$

Moreover, $B^{(i)}_{P_X}(\delta)$ in (15) and $D^{(i)}_{P_{X_{k,b}}}(\delta)$ in (42) are related via

$$B^{(i)}_{P_X}(\delta) = D^{(i)}_{P_{X_{k,b}}}(\delta) + D^{(i)}_{P_{X_{k,\bar{b}}}}(\delta). \tag{43}$$

In analogy to (17), for a uniform input distribution ($p_i = 1/M$)

$$D_{P_X^{\mathrm{u}},\boldsymbol{l}} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}. \tag{44}$$

C. Asymptotic Characterization of BICM

The BICM-GMI is an achievable rate for BICM [16] and is one of the key information-theoretic quantities used to analyze BICM systems. For any P_X and $\boldsymbol{l}$, the BICM-GMI is defined as[4] [22, eq. (24)]

$$\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \sum_{k=1}^{m} \left(\mathsf{I}_{P_X}(\rho) - \sum_{b\in\mathcal{B}} P_{Q_k}(b) \mathsf{I}_{P_{X_{k,b}}}(\rho) \right) \tag{45}$$

and twice its derivative as[5]

$$\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho) \triangleq 2\frac{\mathrm{d}\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{\mathrm{d}\rho} \tag{46}$$

$$= \sum_{k=1}^{m} \left(\mathsf{M}_{P_X}(\rho) - \sum_{b\in\mathcal{B}} P_{Q_k}(b) \mathsf{M}_{P_{X_{k,b}}}(\rho) \right). \tag{47}$$

In these expressions, $\mathsf{I}_{P_{X_{k,b}}}(\rho)$ and $\mathsf{M}_{P_{X_{k,b}}}(\rho)$ are defined, in analogy to (6)–(7), as

$$\mathsf{I}_{P_{X_{k,b}}}(\rho) \triangleq \mathbb{E}_{X,Y}\left[\log\left(f_{Y|X}(Y|X)/f_Y(Y) \right) \right] \tag{48}$$

and

$$\mathsf{M}_{P_{X_{k,b}}}(\rho) \triangleq \mathbb{E}_{X,Y}[(X - \hat{X}^{\mathrm{ME}}_{P_{X_{k,b}}}(Y))^2], \tag{49}$$

$$\hat{X}^{\mathrm{ME}}_{P_{X_{k,b}}}(y) \triangleq \mathbb{E}_X[X|Y=y] \tag{50}$$

where X now follows the distribution $P_{X_{k,b}}$ and Y is the random variable resulting from transmitting $X \in \mathcal{X}_{k,b}$ over the AWGN channel (1). With these definitions, a relation corresponding to (18) between the MI and the MMSE holds also for $\mathsf{I}_{P_{X_{k,b}}}(\rho)$ and $\mathsf{M}_{P_{X_{k,b}}}(\rho)$, and so do theorems analogous to Theorems 1–2, which will be used in the proofs of Theorems 4–5.

Like the MI, the BICM-GMI also tends to H_{P_X} as ρ tends to infinity. The following theorem shows how fast $\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)$ converges to H_{P_X}.

Theorem 4: For any P_X and $\boldsymbol{l}$

$$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \pi D_{P_X,\boldsymbol{l}}. \tag{51}$$

where $D_{P_X,\boldsymbol{l}}$ is given in (41).

Proof: The proof is given in Appendix D. □

The following theorem characterizes the asymptotic behavior of $\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)$.

[4]Even though the BICM-GMI is fully determined by the bit probabilities $[P_{Q_1}(0), P_{Q_2}(0), \ldots, P_{Q_m}(0)]$, we express it as a function of the input distribution P_X in (33).

[5]Since the BICM-GMI is not an MI, its derivative is not an MMSE [31]. We thus avoid using the name MMSE, although we do use an MMSE-like notation $\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)$.

Theorem 5: For any P_X and $\boldsymbol{\imath}$

$$\lim_{\rho\to\infty} \frac{\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} D_{P_X,\boldsymbol{\imath}} \tag{52}$$

where $D_{P_X,\boldsymbol{\imath}}$ is given in (41).

Proof: By using (47) and Theorem 2 we obtain

$$\lim_{\rho\to\infty} \frac{\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} \sum_{k=1}^{m} \left(B_{P_X} - \sum_{b\in\mathcal{B}} P_{Q_k}(b) B_{P_{X_{k,b}}} \right)$$

which in view of Lemma 8 in Appendix D completes the proof. □

In analogy to (8)–(9), we define the BEP as[6]

$$\mathsf{B}_{P_X,\boldsymbol{\imath}}(\rho) \triangleq \frac{1}{m} \sum_{k=1}^{m} \Pr\{\hat{Q}_k^{\mathrm{MAP}}(Y) \neq Q_k\} \tag{53}$$

where Q_k is the transmitted bit and $\hat{Q}_k^{\mathrm{MAP}}(Y)$ is a hard-decision on the bit, i.e., $[\hat{Q}_1^{\mathrm{MAP}}(y), \ldots, \hat{Q}_m^{\mathrm{MAP}}(y)] = \Phi^{-1}(\hat{X}^{\mathrm{MAP}}(y))$ with $\hat{X}^{\mathrm{MAP}}(y)$ given by (9).[7] The next theorem characterizes the asymptotic behavior of the BEP in (53).

Theorem 6: For any P_X and $\boldsymbol{\imath}$

$$\lim_{\rho\to\infty} \frac{\mathsf{B}_{P_X,\boldsymbol{\imath}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{D_{P_X,\boldsymbol{\imath}}}{m} \tag{54}$$

where $D_{P_X,\boldsymbol{\imath}}$ is given in (41).

Proof: The proof is given in Appendix E. □

Similarly to (22)–(24), we can use Theorems 4–6 to show that, at high SNR, the BICM-GMI, twice its derivative, and the BEP behave as

$$\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho) \approx H_{P_X} - \pi D_{P_X,\boldsymbol{\imath}} \mathsf{Q}\left(\frac{\sqrt{\rho}d}{2}\right), \tag{55}$$

$$\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho) \approx \frac{\pi d^2}{4} D_{P_X,\boldsymbol{\imath}} \mathsf{Q}\left(\frac{\sqrt{\rho}d}{2}\right), \tag{56}$$

$$\mathsf{B}_{P_X,\boldsymbol{\imath}}(\rho) \approx \frac{D_{P_X,\boldsymbol{\imath}}}{m} \mathsf{Q}\left(\frac{\sqrt{\rho}d}{2}\right). \tag{57}$$

Thus, at high SNR, the BICM-GMI, twice its derivative, and the BEP have the same asymptotic behavior.[8]

Example 4: Consider the constellation in Example 1, i.e., $\mathcal{X} = \{\pm 4, \pm 2\}$, corresponding to the constituent 4-ary constellation for the unequally spaced 16-ary quadrature amplitude modulation (QAM) constellation specified in the DVB standard [34, Fig. 9b]. Furthermore, we consider the labeling $\boldsymbol{\imath}_{\mathrm{GC}} = [0, 1, 3, 2]$, which gives $A_{\mathcal{X}} = C_{\mathcal{X},\boldsymbol{\imath}} = 4$, and the three input distributions

$$\begin{aligned} \boldsymbol{p}' &= [1/4, 1/4, 1/4, 1/4], \\ \boldsymbol{p}'' &= [1/8, 3/8, 3/8, 1/8], \\ \boldsymbol{p}''' &= [16/25, 4/25, 1/25, 4/25] \end{aligned}$$

[6]Note that (53) is the BEP averaged over the m bit positions, in contrast to the BICM-GMI in (45), which is a sum of m bit-wise MIs.

[7]The BEP in (53) is based on hard-decisions made by the *symbol-wise* MAP demapper. Alternatively, one could study a *bit-wise* MAP demapper for which $\hat{Q}_k^{\mathrm{MAP}}(y) = \operatorname{argmax}_{b\in\mathcal{B}} P_{Q_k|Y}(b|y)$. This demapper is the optimal in terms of BEP, which was recently studied in [32] (see also [33]), but its analysis is much more involved.

[8]Disregarding the "offset" H_{P_X} in (55).

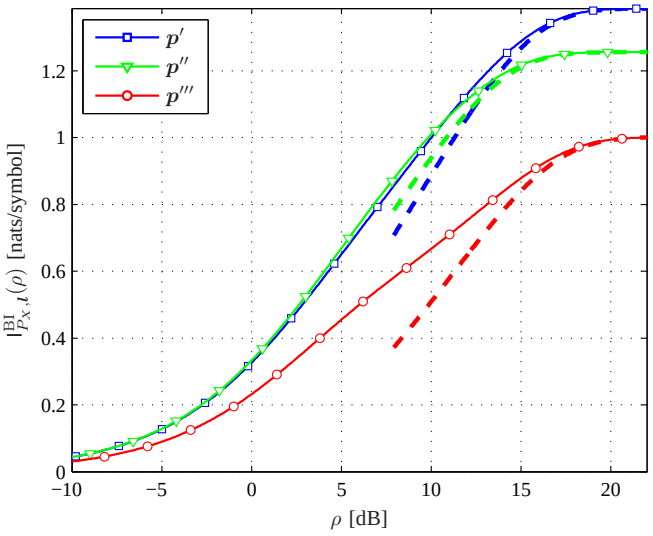

Fig. 5. $\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho)$ for the three input distributions in Example 4 and the constellation $\mathcal{X} = \{\pm 4, \pm 2\}$ (normalized to $E_{\mathrm{s}} = 1$) (solid lines with markers) and the asymptotic expression in (55) (thick dashed lines).

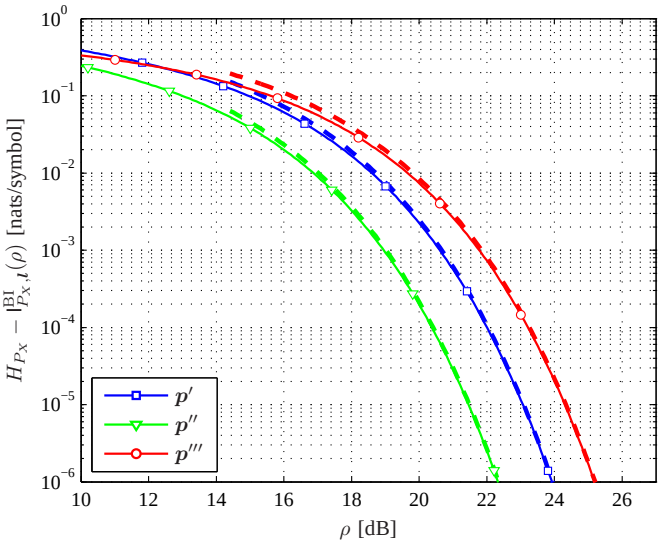

Fig. 6. $H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho)$ for the three input distributions in Example 4 and the constellation $\mathcal{X} = \{\pm 4, \pm 2\}$ (normalized to $E_{\mathrm{s}} = 1$) (solid lines with markers) and the asymptotic expression in (55) (thick dashed lines), i.e., $H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho) \approx \pi D_{P_X,\boldsymbol{\imath}} \mathsf{Q}\left(\sqrt{\rho}d/2\right)$.

which are induced by the bit probabilities listed in the second column of Table II. Table II also lists H_{P_X}, $D_{P_X,\boldsymbol{\imath}}$, and d (when the constellation is normalized to $E_{\mathrm{s}} = 1$). The BICM-GMI curves are shown in Fig. 5. Observe that the BICM-GMI for high SNR converges towards its maximum H_{P_X} (see Table II). The corresponding curves for $H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\imath}}(\rho)$ are shown in Fig. 6. These figures show how the coefficient $D_{P_X,\boldsymbol{\imath}}$ in the asymptotic expression captures the high-SNR behavior of the BICM-GMI for different input distributions.

For a uniform input distribution, Theorems 4–6 particularize to the following result.

Corollary 2: For any $\mathcal{X}$ and $\boldsymbol{\imath}$ and a uniform input distri-

TABLE II
DIFFERENT PARAMETERS FOR THE CONSTELLATION AND INPUT DISTRIBUTIONS IN EXAMPLE 4.

$\boldsymbol{p}$	$P_{Q_1}(0), P_{Q_2}(0)$	H_{P_X}	$D_{P_X,\boldsymbol{l}}$	d
$\boldsymbol{p}'$	$1/2, 1/2$	1.3863	1.0000	0.6325
$\boldsymbol{p}''$	$1/2, 1/4$	1.2555	0.8660	0.7559
$\boldsymbol{p}'''$	$4/5, 4/5$	1.0008	0.8000	0.5423

TABLE III
SUMMARY OF ASYMPTOTICS OF THE BICM-GMI, TWICE ITS DERIVATIVE, AND THE BEP.

Input Distribution	P_X	P_X^{u}
$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	$\pi D_{P_X,\boldsymbol{l}}$	$\pi \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}$
$\lim_{\rho\to\infty} \frac{\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	$\frac{\pi d^2}{4} D_{P_X,\boldsymbol{l}}$	$\frac{\pi d^2}{4} \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}$
$\lim_{\rho\to\infty} \frac{\mathsf{B}_{P_X,\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$	$\frac{D_{P_X,\boldsymbol{l}}}{m}$	$\frac{C_{\mathcal{X},\boldsymbol{l}}}{mM}$

bution

$$\lim_{\rho\to\infty} \frac{\log M - \mathsf{I}^{\mathrm{BI}}_{P_X^{\mathrm{u}},\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \pi \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}, \tag{58}$$

$$\lim_{\rho\to\infty} \frac{\mathsf{M}^{\mathrm{BI}}_{P_X^{\mathrm{u}},\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}, \tag{59}$$

$$\lim_{\rho\to\infty} \frac{\mathsf{B}_{P_X^{\mathrm{u}},\boldsymbol{l}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{mM} \tag{60}$$

where $C_{\mathcal{X},\boldsymbol{l}}$ is given in (37).

Proof: From Theorems 4–6 and (44). □

The expression in (60) corresponds to the well-known expression for the BEP [35, p. 130]. The results in Corollary 2 indicate that, for a uniform input distribution, a maximization of the BICM-GMI is asymptotically equivalent to a minimization of both its derivative and the BEP. The asymptotic results for BICM are summarized in Table III.

D. Lower and Upper Bounds

To study the asymptotic behavior of the BICM-GMI for different labelings $\boldsymbol{l}$, we introduce the two functions

$$\mathsf{K}^{\mathsf{I}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \frac{H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{H_{P_X} - \mathsf{I}_{P_X}(\rho)} \tag{61}$$

$$\mathsf{K}^{\mathsf{M}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \frac{\mathsf{M}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho)}{\mathsf{M}_{P_X}(\rho)}. \tag{62}$$

Noting that $\mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{l}}(\rho) \le \mathsf{I}_{P_X}(\rho)$ [14, eq. (16)], [21, Theorem 5], we have

$$\mathsf{K}^{\mathsf{I}}_{P_X,\boldsymbol{l}}(\rho) \ge 1. \tag{63}$$

We further define

$$\mathsf{R}_{P_X,\boldsymbol{l}} \triangleq \lim_{\rho\to\infty} \mathsf{K}^{\mathsf{I}}_{P_X,\boldsymbol{l}}(\rho) \tag{64}$$

$$= \lim_{\rho\to\infty} \mathsf{K}^{\mathsf{M}}_{P_X,\boldsymbol{l}}(\rho) \tag{65}$$

where (65) follows from L'Hôpital's rule. Theorems 1 and 4 yield

$$\mathsf{R}_{P_X,\boldsymbol{l}} = \frac{D_{P_X,\boldsymbol{l}}}{B_{P_X}} \tag{66}$$

and due to (63)

$$\mathsf{R}_{P_X,\boldsymbol{l}} \ge 1. \tag{67}$$

In the rest of this section, we study $\mathsf{R}_{P_X,\boldsymbol{l}}$ in (66) for a uniform input distribution P_X^{u}. With a slight abuse of notation, we will refer to $\mathsf{R}_{P_X^{\mathrm{u}},\boldsymbol{l}}$ as $\mathsf{R}_{\mathcal{X},\boldsymbol{l}}$.

Lemma 3: For any labeling $\boldsymbol{l}$ and constellation $\mathcal{X}$, $\mathsf{R}_{\mathcal{X},\boldsymbol{l}}$ is given by

$$\mathsf{R}_{\mathcal{X},\boldsymbol{l}} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{A_{\mathcal{X}}}. \tag{68}$$

Proof: Follows by using (44) and (17) in (66). □

Based on Lemma 3, an upper bound on $\mathsf{R}_{\mathcal{X},\boldsymbol{l}}$ can be obtained as follows.

Theorem 7: For any one-dimensional constellation and any labeling $\boldsymbol{l}$

$$C_{\mathcal{X},\boldsymbol{l}} \le \min\left(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M\right) \tag{69}$$

and thus,

$$\mathsf{R}_{\mathcal{X},\boldsymbol{l}} \le \frac{\min\left(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M\right)}{A_{\mathcal{X}}}. \tag{70}$$

Proof: We note that for any labeling there are exactly $M/2$ pairs of labels at Hamming distance m. Because of this, at most $M/2$ pairs of constellation points at MED can each differ in exactly m bits, which can be the case only if $A_{\mathcal{X}} \le M$. This case gives $C_{\mathcal{X},\boldsymbol{l}} \le mA_{\mathcal{X}}$. If there are more than $M/2$ pairs of constellation points at MED, i.e., $A_{\mathcal{X}} > M$, $M/2$ pairs can differ in m bits and the remaining $(A_{\mathcal{X}} - M)/2$ pairs can differ in at most $m-1$ bits, which gives $C_{\mathcal{X},\boldsymbol{l}} \le mM + (m-1)(A_{\mathcal{X}} - M) = (m-1)A_{\mathcal{X}} + M$. The expression in (70) follows from (69) and (68). □

For an MPAM constellation, using (14), Lemma 3 and Theorem 7 specialize into

$$\mathsf{R}_{\mathcal{E},\boldsymbol{l}} = \frac{C_{\mathcal{E},\boldsymbol{l}}}{2(M-1)}, \tag{71}$$

$$\mathsf{R}_{\mathcal{E},\boldsymbol{l}} \le m - \frac{M-2}{2M-2}. \tag{72}$$

Furthermore, if the MPAM constellation is labeled with the NBC, we obtain via (40)

$$\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} = \frac{2M - m - 2}{M-1}. \tag{73}$$

Example 5: In Fig. 7, we show the functions $\mathsf{K}^{\mathsf{I}}_{P_X^{\mathrm{eu}},\boldsymbol{l}}(\rho)$ and $\mathsf{K}^{\mathsf{M}}_{P_X^{\mathrm{eu}},\boldsymbol{l}}(\rho)$ in (61) and (62), respectively, for a 4PAM constellation with a uniform input distribution ($P_X = P_X^{\mathrm{eu}}$, $A_{\mathcal{X}} = 6$) and the three labelings that give different BICM-GMI: $\boldsymbol{l}_{\mathrm{GC}} = [0,1,3,2]$, $\boldsymbol{l}_{\mathrm{NBC}} = [0,1,2,3]$, and $\boldsymbol{l}_{\mathrm{AGC}} =$

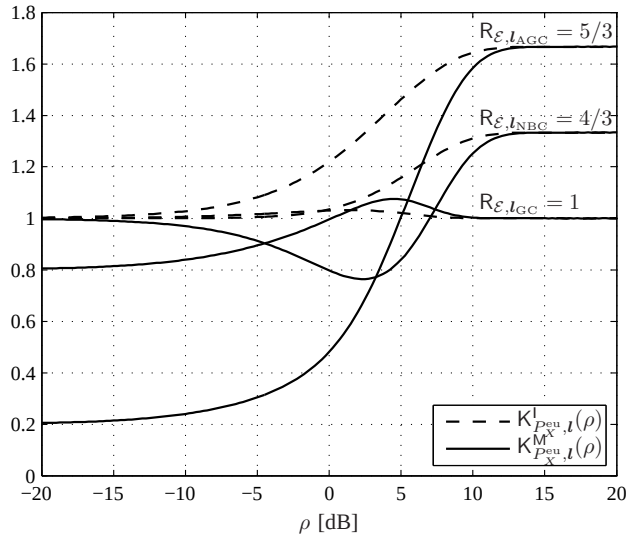

Fig. 7. Functions $\mathsf{K}^{\mathsf{I}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ (dashed lines) and $\mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ (solid lines) for 4PAM (normalized to $E_{\mathrm{s}} = 1$) and all three nonequivalent labelings. The values of $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$ in (71) are also shown.

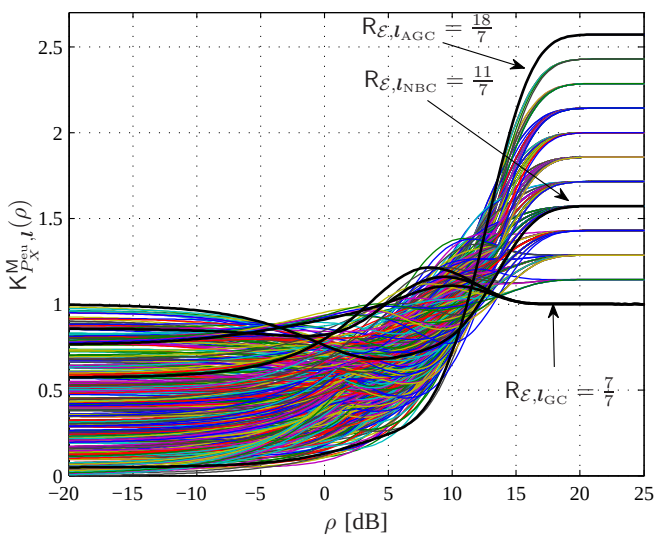

Fig. 8. Function $\mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ for the 458 labelings that give a different BICM-GMI for 8PAM (normalized to $E_{\mathrm{s}} = 1$). The values of $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$ in (71) for the three nonequivalent GCs, the NBC, and the AGC are also shown.

$[0, 3, 2, 1]$.[9] The values of $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$ in (71) are also shown. In contrast to the BICM-GMI curves plotted, e.g., in [17, Fig. 3] and [31, Fig. 1], the functions $\mathsf{K}^{\mathsf{I}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ and $\mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ allow us to study different labelings at high SNR. Observe that the GC (i.e., $\boldsymbol{l}_{\mathrm{GC}}$), gives $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{GC}}} = 1$, and that $\boldsymbol{l}_{\mathrm{AGC}}$ achieves the upper bound in (72), i.e., $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{AGC}}} = 5/3$.

The function $\mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ also allows us to study different labelings at low SNR: Fig. 7 shows that the NBC is the binary labeling for 4PAM that gives the largest value for $\mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ as ρ tends to zero, which agrees with [20], [21, Theorem 14].[10] Recall that the best labeling in terms of $\mathsf{I}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ at low SNR is by (18) the worst one in terms of $\mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ (i.e., the one that maximizes $\mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$). Furthermore, a labeling that gives a high $\mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ at low SNR tends to yield a low $\mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ at high SNR, since

$$\int_0^\infty \mathsf{M}^{\mathrm{BI}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)d\rho = 2\log M \tag{74}$$

is constant for a given constellation.

Example 6: In Fig. 8, we show the function $\mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ for 8PAM ($P_X = P^{\mathrm{eu}}_X$, $A_{\mathcal{X}} = 14$) and all the 458 labelings that give a different BICM-GMI [23]. In this figure, 12 possible values of $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$ in (71) are clearly visible, which coincide with the results in [23, Fig. 3].[11] Using (60), the 12 values of $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$ in Fig. 8 also translate into 12 different asymptotic BEP curves, which were recently reported in [33, Fig. 4]. The value $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}}$ obtained using (73) is also shown. A careful examination of Fig. 8 reveals that there are three labelings minimizing $\mathsf{R}_{\mathcal{E},\boldsymbol{l}}$. These are the three nonequivalent GCs (in terms of BEP) [28, Table I]: the BRGC $\boldsymbol{l} = [0, 1, 3, 2, 6, 7, 5, 4]$, $\boldsymbol{l} = [0, 1, 3, 2, 6, 4, 5, 7]$, and $\boldsymbol{l} = [0, 1, 3, 7, 5, 4, 6, 2]$.

[9]The anti-Gray code (AGC) will be formally introduced in Sec. IV-E.

[10]The relationship between the coefficient α determining the low-SNR behavior of a zero-mean constellation with a uniform input distribution [21, eq. (47)] is $\alpha \log 2 = \lim_{\rho\to 0} \mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ (see also [24, eq. (86)]).

[11]Further note that $\lim_{\rho\to 0} \mathsf{K}^{\mathsf{M}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ reveals the 72 classes of labelings reported in [21, Fig. 6 (a)].

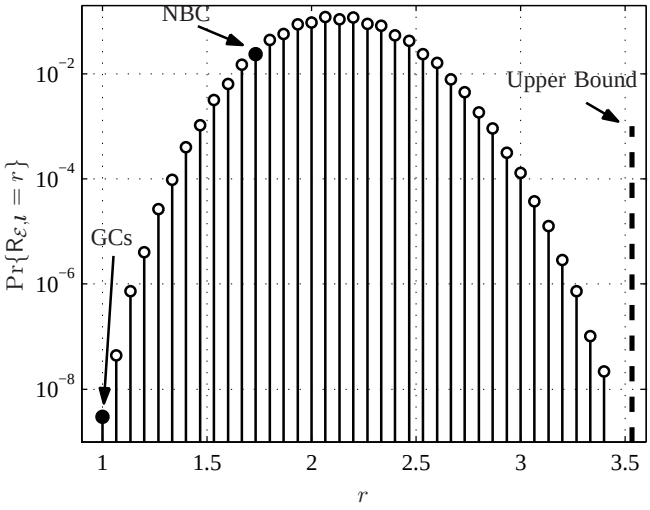

Fig. 9. Approximated $\Pr\{\mathsf{R}_{\mathcal{E},\boldsymbol{l}} = r\}$ using 10^9 randomly generated labelings for 16PAM (normalized to $E_{\mathrm{s}} = 1$). For GCs $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{GC}}} = 1$ and for the NBC $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} = 26/15$. The upper bound in (72) is also shown.

Example 7: Motivated by [21, Fig. 6], we present in Fig. 9 an approximation for the PMF $\Pr\{\mathsf{R}_{\mathcal{X},\boldsymbol{l}} = r\}$ for 16PAM obtained by randomly generating 10^9 labelings. This figure shows that most of the possible labelings are not Gray. For $M = 16$, we obtain $\mathsf{R}_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} = 26/15$, see (73), which is highlighted in Fig. 9. The upper bound in (72) is also shown. In the next section, we will show how to construct a labeling that achieves this upper bound.

E. Gray Codes and Anti-Gray Codes

In view of the lower bound (67), we say that, for a constellation $\mathcal{X}$ and a uniform input distribution, a labeling $\boldsymbol{l}$ is *asymptotically optimal* (AO) in terms of BICM-GMI if

it satisfies $\mathsf{R}_{\mathcal{X},\boldsymbol{\imath}} = 1$. Intuitively, an AO labeling is a binary labeling for which the BICM-GMI approaches H_{P_X} as fast as the MI does for the same constellation $\mathcal{X}$.

By inspection of (73), we see that the NBC for MPAM is not an AO labeling for $m \geq 2$. The following theorem demonstrates that GCs are AO at high SNR. Thus, it proves a special case of the conjecture of the optimality of GCs at high SNR in terms of BICM-GMI [14, Sec. III-C]. (It has previously been disproved for low to medium SNRs [17].)

Theorem 8: For any constellation $\mathcal{X}$ and a uniform input distribution, a labeling is AO if and only if it is a GC.

Proof: For any GC, all pairs of constellation points at MED are at Hamming distance one. Thus, (39) holds with equality, and by (68), $\mathsf{R}_{\mathcal{X},\boldsymbol{\imath}} = 1$. This completes the "if" part of the proof. The "only if" part follows because for any non-GC, there is at least one pair of constellation points at Hamming distance larger than one, thus, $C_{\mathcal{X},\boldsymbol{\imath}} > A_{\mathcal{X}}$, and therefore, $\mathsf{R}_{\mathcal{X},\boldsymbol{\imath}} > 1$. □

Remark 3: The results about the optimality of GCs directly extend to multidimensional constellations that are constructed as direct products of one-dimensional constellations, provided that the labeling is generated via an ordered direct product of GCs. This construction of constellation and labelings was formally used, e.g., in [21, Theorem 15].

Remark 4: While the NBC is not AO for an MPAM constellation, it may be AO for an unequally spaced constellation. For example, this is the case if the NBC is used with the constellation in Example 1, in which case the NBC is a GC, according to the definition in Section IV-B.

Theorem 8 shows that GCs minimize $\mathsf{R}_{\mathcal{X},\boldsymbol{\imath}}$. In what follows, we show that, for MPAM constellations, it is always possible to construct a labeling that maximizes $\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}}$, i.e., a labeling that achieves the upper bound in (72).

We define the set of all possible values that $C_{\mathcal{X},\boldsymbol{\imath}}$ can take as $\mathcal{C}_{\mathcal{X}}$, where

$$|\mathcal{C}_{\mathcal{X}}| \leq \frac{1}{2} \min\left\{(m-1)A_{\mathcal{X}} + 2, (m-2)A_{\mathcal{X}} + M + 2\right\}. \tag{75}$$

This inequality follows because $C_{\mathcal{X},\boldsymbol{\imath}}$ is an even integer bounded by (39) and (69).

The expression (75) is an upper bound on the number of classes of labelings with different high-SNR behavior in terms of BICM-GMI (or equivalently BEP). For the particular case of $\mathcal{X} = \mathcal{E}$, by using (14) in (75), we obtain

$$|\mathcal{C}_{\mathcal{E}}| \leq mM - \frac{3M}{2} - m + 3. \tag{76}$$

For 4PAM we have $|\mathcal{C}_{\mathcal{E}}| \leq 3$ and for 8PAM we have $|\mathcal{C}_{\mathcal{E}}| \leq 12$, which coincides with the 3 and 12 classes at high SNR shown in Fig. 7 and Fig. 8, respectively. For 16PAM, the upper bound (76) indicates that $|\mathcal{C}_{\mathcal{E}}| \leq 39$. However, Fig. 9 shows only 37 classes. This raises the question of the tightness of the bound in (76) (or equivalently, the upper bound in (72)), which we address in the following.

The AGC of order $m \geq 2$ is defined by the $M \times m$ binary matrix $\mathbf{W}_m$, where the ith row is the binary label for x_i, where $\mathbf{W}_1 = [0, 1]^{\mathsf{T}}$, and where the following steps construct $\mathbf{W}_m$ from $\mathbf{W}_{m-1}$:

Step 1 Reverse the $M/2$ rows in $\mathbf{W}_{m-1}$, and append them below $\mathbf{W}_{m-1}$ to construct a new matrix $\mathbf{W}'_m$ with M rows and $m-1$ columns.

Step 2 Append the length M column vector $[0, 1, 0, 1, \ldots, 0, 1]^{\mathsf{T}}$ to the left of $\mathbf{W}'_m$ to create $\mathbf{W}''_m$, with M rows and m columns.

Step 3 Negate all bits in the lower half of $\mathbf{W}''_m$ to obtain $\mathbf{W}_m$.

The recursive construction described above is illustrated in Fig. 10 for $m = 2$ and $m = 3$. The following lemma shows that this construction indeed leads to a valid labeling.

Lemma 4: All the rows in $\mathbf{W}_m$ are unique, and thus, the AGC is a valid labeling.

Proof: Consider the above construction of an AGC. Assume that $\mathbf{W}_{m-1}$ is a valid labeling (all rows are unique) where every odd row differs in $m-1$ bits compared to the row below. $\mathbf{W}_1$ fulfills both criteria, since it is a valid labeling where the first row differs in 1 bit compared to the second row. Because of Step 1, every odd row in the upper half of $\mathbf{W}'_m$ is identical to an even row in the lower half of $\mathbf{W}'_m$, which directly implies that all rows of $\mathbf{W}''_m$ in Step 2 are unique. Thus, $\mathbf{W}''_m$ is a valid labeling. It also implies that every odd row of $\mathbf{W}''_m$ differs in m bits compared to the row below, since the corresponding rows of $\mathbf{W}'_m$ differ in $m-1$ bits. Inverting all the bits in the lower half of $\mathbf{W}''_m$ is therefore equivalent to swapping every odd row in the lower half of $\mathbf{W}''_m$ with the row below. This operation makes $\mathbf{W}_m$ a valid labeling with M unique rows, where every odd row differs in m bits compared to the row below. □

The next theorem proves that, at high SNR, the AGC is the worst binary labeling for MPAM constellations.

Theorem 9: For $\mathcal{X} = \mathcal{E}$, the AGC achieves the upper bound in (72), i.e.,

$$\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}_{\mathrm{AGC}}} = m - \frac{M-2}{2M-2}. \tag{77}$$

Proof: Let $H_m = C_{\mathcal{E},\boldsymbol{\imath}_{\mathrm{AGC}}}$ denote twice the sum of the Hamming distances between all adjacent rows in $\mathbf{W}_m$, and let H'_m and H''_m denote the same quantity for $\mathbf{W}'_m$ and $\mathbf{W}''_m$, respectively. Steps 1 and 2 give $H'_m = 2H_{m-1}$ and $H''_m = H'_m + 2(M-1)$. It then follows that $H_m = H''_m - 2 + 2(m-1)$, since row $M/2$ and row $M/2+1$ in $\mathbf{W}''_m$ differ in only one bit and therefore the same rows in $\mathbf{W}_m$ differ in $m-1$ bits. This gives $H_m = 2H_{m-1} + 2(M+m-3)$, which combined with $H_1 = 2$ gives $H_m = 2\left(mM - m - M/2 + 1\right)$. Together with (71), this completes the proof. □

The labeling $\boldsymbol{\imath}_3$ in Example 5 and Fig. 7 (i.e., $\mathbf{W}_2$ in Fig. 10) is the AGC for 4PAM with $\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}} = 5/3$ given by (77). For 8PAM, the AGC is $\boldsymbol{\imath}_{\mathrm{AGC}} = [0, 7, 2, 5, 6, 1, 4, 3]$ ($\mathbf{W}_3$ in Fig. 10), whose corresponding function $\mathsf{K}^{!}_{P_X^{\mathrm{eu}},\boldsymbol{\imath}_{\mathrm{AGC}}}(\rho)$ is shown in Fig. 8, with $\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}} = 18/7$.

For $M = 16$, the labeling that maximizes $\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}}$ ($\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}} = 106/30 \approx 3.53$) is the AGC $\mathbf{W}_4$ (as shown by Theorem 9), which can be constructed as described before. It can be further shown that the labeling with the second largest $\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}}$ ($\mathsf{R}_{\mathcal{E},\boldsymbol{\imath}} = 104/30 \approx 3.47$) can be constructed by reversing the order of the three first rows of the AGC $\mathbf{W}_4$. This demonstrates that for 16PAM all 39 classes are indeed possible. The last two classes are not shown in Fig. 9 because the total number of labelings in this case is $16! \approx 2.1 \cdot 10^{13}$ (without discarding

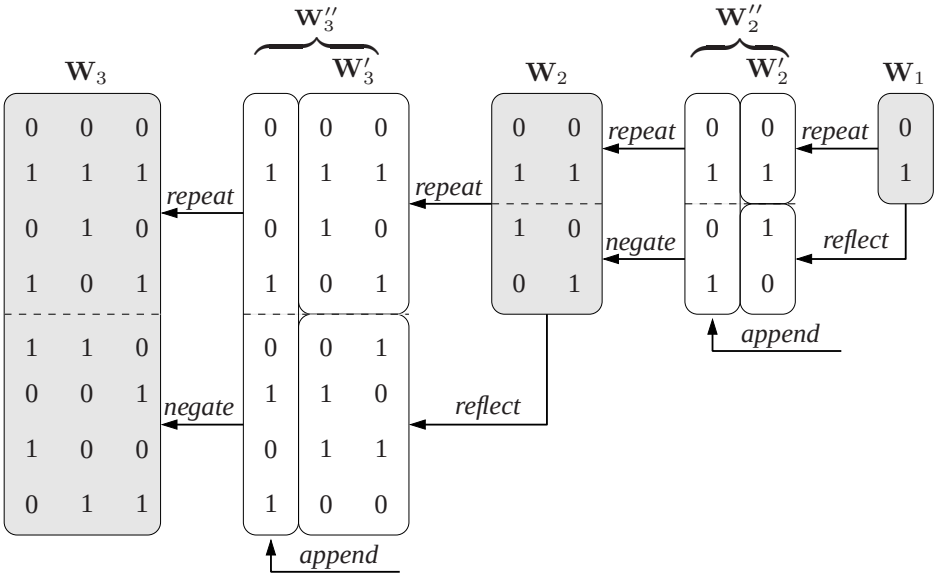

Fig. 10. Proposed recursive construction of an AGC for $m = 2$ and $m = 3$.

trivial operations), so randomly generating 10^9 labelings only covers a small fraction of all possible labelings.

V. Conclusions

In this paper, we studied discrete constellations with arbitrary input distributions over the scalar AWGN channel in the high-SNR regime and derived exact asymptotic expressions for key quantities in information theory, estimation theory, and communication theory: the MI, MMSE, SEP, the BICM-GMI, its derivative, and BEP. Our results show that, as the SNR tends to infinity, all these quantities converge to their asymptotes proportionally to $\mathsf{Q}\left(\sqrt{\rho}d/2\right)$, where d is the MED of the constellation. This demonstrates the asymptotic equivalence between all these quantities as well as the importance of the Gaussian Q-function.

For a uniform input distribution, the proportionality constants for the MI, SEP, and MMSE were found to be a function of the MED of the constellation and the number of pairs of constellation points at MED only, and thus, the constellation that maximizes the MI in the high-SNR regime is the same that minimizes both the SEP and the MMSE.

We then applied our results to the problem of binary labelings for BICM. By characterizing the high-SNR behavior of the BICM-GMI, asymptotically optimal binary labelings were found, and the long-standing conjecture that Gray codes are optimal at high SNR was proved. We also proved that there always exists an anti-Gray code for MPAM constellations, which is the labeling that has the lowest BICM-GMI and the highest BEP at high SNR.

Appendix A
Proof of Theorem 1

We start by upper and lower bounding the Q-function via [36, Prop. 19.4.2]

$$\left(1-\frac{1}{x^2}\right)G(x) \le \mathsf{Q}(x) \le G(x), \quad x>0 \tag{78}$$

where

$$G(x) \triangleq \frac{1}{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}. \tag{79}$$

It follows that

$$\lim_{x\to\infty}\frac{G(x)}{\mathsf{Q}(x)} = 1. \tag{80}$$

The MI in (6) can be expressed as

$$\begin{aligned}\mathsf{I}_{P_X}(\rho) = \sum_{i\in\mathcal{I}_\mathcal{X}} p_i \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_i)^2} \\ \cdot \log\frac{\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_i)^2}}{\sum_{j\in\mathcal{I}_\mathcal{X}} p_j \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}\,\mathrm{d}y \end{aligned} \tag{81}$$

$$= -\sum_{i\in\mathcal{I}_\mathcal{X}} p_i \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{\sqrt{\pi}} \log \sum_{\delta\in\mathcal{D}_\mathcal{X}^{(i)}} p_j \mathrm{e}^{-\sqrt{2\rho}t\delta-\frac{\rho\delta^2}{2}}\,\mathrm{d}t \tag{82}$$

where to pass from (81) to (82) we used the substitution $y - \sqrt{\rho}x_i = \sqrt{2}t$ and

$$\mathcal{D}_\mathcal{X}^{(i)} \triangleq \{x_i - x : x\in\mathcal{X}\}. \tag{83}$$

Using (82) and the definition of entropy, the numerator of the left-hand side (l.h.s.) of (19) can be expressed as

$$H_{P_X} - \mathsf{I}_{P_X}(\rho) = \sum_{i\in\mathcal{I}_\mathcal{X}} p_i V_i(\rho) \tag{84}$$

where

$$V_i(\rho) \triangleq \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{\sqrt{\pi}} \log \sum_{\delta\in\mathcal{D}_\mathcal{X}^{(i)}} R_{P_X}^{(i)}(\delta)\cdot \mathrm{e}^{-\sqrt{2\rho}t\delta-\frac{\rho\delta^2}{2}}\,\mathrm{d}t \tag{85}$$

and

$$R_{P_X}^{(i)}(\delta) \triangleq \left(\frac{B_{P_X}^{(i)}(\delta)}{p_i}\right)^2 \tag{86}$$

$$= \begin{cases} \frac{p_j}{p_i}, & \text{if } \exists x_j \in \mathcal{X} : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases}. \tag{87}$$

Combining (80) and (84) yields

$$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{G\left(\sqrt{\rho}d/2\right)} \tag{88}$$

$$= \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \lim_{\rho\to\infty} \frac{V_i(\rho)}{G\left(\sqrt{\rho}d/2\right)}. \tag{89}$$

As will become apparent later, the limit on the right-hand side (r.h.s.) of (89) exists and, hence, so does the limit on the l.h.s. of (88).

In what follows, we calculate the limit on the r.h.s. of (89). Using (79) and (85), and substituting $r = d\sqrt{\rho/8}$, we obtain

$$\lim_{\rho\to\infty} \frac{V_i(\rho)}{G\left(\sqrt{\rho}d/2\right)} = 2\left(\lim_{r\to\infty} F_i^-(r) + \lim_{r\to\infty} F_i^+(r)\right) \tag{90}$$

where

$$F_i^-(r) \triangleq \int_{-\infty}^{0} r\mathrm{e}^{r^2-t^2} \log \sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta)\cdot \mathrm{e}^{-4rt\frac{\delta}{d}-4r^2\frac{\delta^2}{d^2}}\,\mathrm{d}t \tag{91}$$

and

$$F_i^+(r) \triangleq \int_{0}^{\infty} r\mathrm{e}^{r^2-t^2} \log \sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta)\cdot \mathrm{e}^{-4rt\frac{\delta}{d}-4r^2\frac{\delta^2}{d^2}}\,\mathrm{d}t. \tag{92}$$

We begin with the first limit on the r.h.s. of (90). Using the substitution $t = u/r - r$, we express $F_i^-(r)$ in (91) as

$$F_i^-(r) = \int_{-\infty}^{r^2} \mathrm{e}^{2u-\frac{u^2}{r^2}} \log \sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta)\cdot \mathrm{e}^{-4u\frac{\delta}{d}-4r^2U(\delta)}\,\mathrm{d}u \tag{93}$$

where

$$U(\delta) \triangleq \frac{\delta}{d}\left(\frac{\delta}{d}-1\right). \tag{94}$$

Note that $U(\delta) \geq 0, \forall \delta \in \mathcal{D}_{\mathcal{X}}$. Defining

$$f_i^-(r,u) \triangleq h(r^2-u)\cdot \mathrm{e}^{2u-\frac{u^2}{r^2}} \cdot \log\left(1 + \sum_{\delta\in\mathcal{D}_i^*} R_{P_X}^{(i)}(\delta)\cdot \mathrm{e}^{-4u\frac{\delta}{d}-4r^2U(\delta)}\right) \tag{95}$$

with $\mathcal{D}_i^* \triangleq \mathcal{D}_{\mathcal{X}}^{(i)}\backslash\{0\}$ and $h(x)$ being Heaviside's step function (i.e., $h(x) = 1$ if $x \geq 0$ and $h(x) = 0$ if $x < 0$), $F_i^-(r)$ in (93) can be written as

$$F_i^-(r) = \int_{-\infty}^{\infty} f_i^-(r,u)\,\mathrm{d}u. \tag{96}$$

Note that, for every $r > 0$, the function $u \mapsto f_i^-(r,u)$ is nonnegative. Further note that $U(d) = 0$,

$$\lim_{r\to\infty} f_i^-(r,u) = \mathrm{e}^{2u}\log\left(1 + R_{P_X}^{(i)}(d)\cdot\mathrm{e}^{-4u}\right),\ u\in\mathbb{R}. \tag{97}$$

We will show that, for every $r > 0$, $u \mapsto f_i^-(r,u)$ is uniformly bounded by some integrable function $u \mapsto g_i^-(u)$ that is independent of r (see Lemma 5 ahead). To compute the first limit on the r.h.s. of (90), we can thus use Lebesgue's Dominated Convergence Theorem [37, Theorem 1.34] to obtain

$$\lim_{r\to\infty} F_i^-(r) = \lim_{r\to\infty} \int_{-\infty}^{\infty} f_i^-(r,u)\,\mathrm{d}u \tag{98}$$

$$= \int_{-\infty}^{\infty} \lim_{r\to\infty} f_i^-(r,u)\,\mathrm{d}u \tag{99}$$

$$= \int_{-\infty}^{\infty} \mathrm{e}^{2u}\log\left(1 + R_{P_X}^{(i)}(d)\cdot\mathrm{e}^{-4u}\right)\,\mathrm{d}u \tag{100}$$

$$= \frac{\pi\sqrt{R_{P_X}^{(i)}(d)}}{2} \tag{101}$$

where (100) is obtained from (97) and (101) follows from the substitution $x^2 = R_{P_X}^{(i)}(d)\mathrm{e}^{-4u}$ together with [38, eq. (4.295.3)].

It thus remains to show that $u \mapsto f_i^-(r,u)$ is uniformly bounded by some integrable function $g_i^-(u)$ that is independent of r. We do this in the following lemma.

Lemma 5: For any $r > 0$

$$0 \leq f_i^-(r,u) \leq g_i^-(u), \quad u\in\mathbb{R} \tag{102}$$

where

$$g_i^-(u) \triangleq \begin{cases} \mathrm{e}^{2u}\log\left(\frac{M}{p_i}\mathrm{e}^{-4u\hat{d}^2/d^2}\right), & u < 0 \\ \mathrm{e}^{2u}\log\left(1 + \frac{M-1}{p_i}\mathrm{e}^{-4u}\right), & u \geq 0 \end{cases} \tag{103}$$

and $\hat{d}$ is the maximum ED of the constellation, i.e., $\hat{d} \triangleq \max_{x_i,x_j\in\mathcal{X}}|x_i - x_j|$. Furthermore,

$$\int_{-\infty}^{\infty} g_i^-(u)\,\mathrm{d}u < \infty. \tag{104}$$

Proof: We first note that, for every $r > 0$, the function $u \mapsto f_i^-(r,u)$ is nonnegative. It thus remains to show the second inequality in (102). To this end, we use $\mathrm{e}^{-\frac{u^2}{r^2}} \leq 1$ and $R_{P_X}^{(i)}(d) < 1/p_i$ to upper-bound (95) as

$$f_i^-(r,u) \leq h(r^2-u)\mathrm{e}^{2u}\log\left(1 + \sum_{\delta\in\mathcal{D}_i^*} \frac{\mathrm{e}^{-4u\frac{\delta}{d}-4r^2U(\delta)}}{p_i}\right) \tag{105}$$

$$\leq \mathrm{e}^{2u}\log\left(1 + \sum_{\delta\in\mathcal{D}_i^*} \frac{\mathrm{e}^{-4u\frac{\delta^2}{d^2}}}{p_i}\right) \tag{106}$$

where to pass from (105) to (106) we used $\mathrm{e}^{-4r^2U(\delta)} \leq \mathrm{e}^{-4uU(\delta)}$ for $u \leq r^2$ (because $U(\delta) \geq 0$) and that the r.h.s. of (106) is nonnegative for $u < r^2$.

For $u \geq 0$, we have

$$f_i^-(r,u) \leq \mathrm{e}^{2u}\log\left(1 + \frac{M-1}{p_i}\mathrm{e}^{-4u}\right) \tag{107}$$

which is obtained by applying $\delta^2 \geq d^2, \delta \in \mathcal{D}_i^*$ in (106). For $u < 0$, (106) is upper-bounded by

$$f_i^-(r,u) \leq \mathrm{e}^{2u} \log \sum_{\delta \in \mathcal{D}_i} \frac{\mathrm{e}^{-4u\delta^2/d^2}}{p_i} \tag{108}$$

$$\leq \mathrm{e}^{2u} \log \left(\frac{M}{p_i} \mathrm{e}^{-4u\hat{d}^2/d^2} \right) \tag{109}$$

where (108) is obtained by using $1 < 1/p_i$ and (109) follows from $\delta^2 \leq d^2, \delta \in \mathcal{D}_i$.

To prove (104), we write

$$\int_{-\infty}^{\infty} g_i^-(u)\,\mathrm{d}u = \int_{-\infty}^{0} g_i^-(u)\,\mathrm{d}u + \int_{0}^{\infty} g_i^-(u)\,\mathrm{d}u \tag{110}$$

where from (109)

$$\int_{-\infty}^{0} g_i^-(u)\,\mathrm{d}u = \frac{1}{2}\log\frac{M}{p_i} + \frac{\hat{d}^2}{d^2} \tag{111}$$

and from (107)

$$\int_{0}^{\infty} g_i^-(u)\,\mathrm{d}u \leq \int_{-\infty}^{\infty} \mathrm{e}^{2u} \log\left(1 + \frac{M-1}{p_i}\mathrm{e}^{-4u}\right)\,\mathrm{d}u$$

$$= \frac{\pi}{2}\sqrt{\frac{M-1}{p_i}} \tag{112}$$

which follows in analogy to (100)–(101). This completes the proof of Lemma 5. □

Returning to the proof of Theorem 1, the second limit on the r.h.s. of (90) can be computed along the same lines by substituting $t = u/r + r$ in (92), which gives

$$\lim_{r\to\infty} F_i^+(r) = \frac{\pi\sqrt{R_{P_X}^{(i)}(-d)}}{2}. \tag{113}$$

Combining (101) and (113) with (90) and (89) yields

$$\lim_{\rho\to\infty} \frac{H_{P_X} - \mathsf{I}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \pi \left(\sqrt{R_{P_X}^{(i)}(d)} + \sqrt{R_{P_X}^{(i)}(-d)} \right) \tag{114}$$

which in view of (86) and (16) is equal to πB_{P_X}. This proves Theorem 1.

Appendix B
Proof of Theorem 2

For the AWGN channel in (1), the conditional mean estimator is given by

$$\hat{X}^{\mathrm{ME}}(y) = \frac{\sum_{j\in\mathcal{I}_{\mathcal{X}}} p_j x_j \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}{\sum_{j\in\mathcal{I}_{\mathcal{X}}} p_j \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}. \tag{115}$$

By using (115) in (7), we obtain

$$\mathsf{M}_{P_X}(\rho) = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_i)^2} \cdot \left(\frac{\sum_{j\in\mathcal{I}_{\mathcal{X}}} p_j (x_i - x_j) \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}{\sum_{j\in\mathcal{I}_{\mathcal{X}}} p_j \mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}} \right)^2 \mathrm{d}y \tag{116}$$

$$= \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \tilde{V}_i(\rho) \tag{117}$$

where

$$\tilde{V}_i(\rho) \triangleq \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-t^2}}{\sqrt{\pi}} \left(\frac{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} \delta R_{P_X}^{(i)}(\delta) \cdot \mathrm{e}^{-\sqrt{2\rho}t\delta - \frac{\rho\delta^2}{2}}}{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta) \cdot \mathrm{e}^{-\sqrt{2\rho}t\delta - \frac{\rho\delta^2}{2}}} \right)^2 \mathrm{d}t \tag{118}$$

and where $R_{P_X}^{(i)}(\delta)$ is given by (87). To pass from (116) to (117) we used the substitution $y - \sqrt{\rho}x_i = \sqrt{2}t$.

Using (117), we obtain

$$\lim_{\rho\to\infty} \frac{\mathsf{M}_{P_X}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \lim_{\rho\to\infty} \frac{\tilde{V}_i(\rho)}{G\left(\sqrt{\rho}d/2\right)}. \tag{119}$$

As will become apparent later, the limit on the r.h.s. of (119) exists and, hence, so does the limit on the l.h.s.. To compute the limit on the r.h.s. of (119), we shall follow similar steps to those in Appendix A. We will therefore omit some intermediate steps.

Using (118), (79) and the substitution $r = d\sqrt{\rho/8}$, we have

$$\lim_{\rho\to\infty} \frac{\tilde{V}_i(\rho)}{G\left(\sqrt{\rho}d/2\right)} = 2\left(\lim_{r\to\infty} \tilde{F}_i^-(r) + \lim_{r\to\infty} \tilde{F}_i^+(r) \right) \tag{120}$$

where

$$\tilde{F}_i^-(r) \triangleq \int_{-\infty}^{0} r\mathrm{e}^{r^2-t^2} \left(\frac{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} \delta R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}}}{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}}} \right)^2 \mathrm{d}t, \tag{121}$$

$$\tilde{F}_i^+(r) \triangleq \int_{0}^{\infty} r\mathrm{e}^{r^2-t^2} \left(\frac{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} \delta R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}}}{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}} R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4rt\frac{\delta}{d} - 4r^2\frac{\delta^2}{d^2}}} \right)^2 \mathrm{d}t. \tag{122}$$

We will now calculate the first limit in (120). Using the substitution $t = u/r - r$ we express $\tilde{F}_i^-(r)$ in (121) as

$$\tilde{F}_i^-(r) = \int_{-\infty}^{\infty} \tilde{f}_i^-(r,u)\,\mathrm{d}u \tag{123}$$

where

$$\tilde{f}_i^-(r,u) \triangleq h(r^2-u) \cdot \mathrm{e}^{2u - \frac{u^2}{r^2}} \cdot \left(\frac{\sum_{\delta\in\mathcal{D}_i^*} \delta R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4u\frac{\delta}{d} - 4r^2 U(\delta)}}{1 + \sum_{\delta\in\mathcal{D}_i^*} R_{P_X}^{(i)}(\delta) \mathrm{e}^{-4u\frac{\delta}{d} - 4r^2 U(\delta)}} \right)^2. \tag{124}$$

Recall that $U(\delta)$ is given by (94) and $h(x)$ is the Heaviside's step function. Using the fact that $U(d) = 0$ and $U(\delta) \geq 0, \forall \delta \in \mathcal{D}_{\mathcal{X}}$, we obtain

$$\lim_{r\to\infty} \tilde{f}_i^-(r,u) = d^2 \mathrm{e}^{2u} \left(\frac{R_{P_X}^{(i)}(d)\mathrm{e}^{-4u}}{1 + R_{P_X}^{(i)}(d)\mathrm{e}^{-4u}} \right)^2. \tag{125}$$

As we shall prove in Lemma 6 ahead, $u \mapsto \tilde{f}_i^-(r,u)$ is uniformly bounded by some integrable function $u \mapsto \tilde{g}_i(u)$ that is independent of r. It thus follows from Lebesgue's

Dominated Convergence Theorem that

$$\lim_{r\to\infty}\tilde{F}_i^-(r)=\int_{-\infty}^{\infty}\lim_{r\to\infty}\tilde{f}_i^-(r,u)\,\mathrm{d}u \tag{126}$$

$$=\frac{d^2\sqrt{R_{P_X}^{(i)}(d)}}{2}\int_0^{\infty}\frac{x^2}{(1+x^2)^2}\,\mathrm{d}x \tag{127}$$

$$=\frac{d^2\pi\sqrt{R_{P_X}^{(i)}(d)}}{8} \tag{128}$$

where (127) follows from (125) and the substitution $\sqrt{R_{P_X}^{(i)}(d)}\mathrm{e}^{-2u}=x$, and (128) follows from [38, eq. (3.241.5)].

The second limit in the r.h.s. of (120) can be computed along the same lines by using the substitution $t=u/r+r$ in (122). We obtain

$$\lim_{r\to\infty}\tilde{F}_i^+(r)=\frac{d^2\pi\sqrt{R_{P_X}^{(i)}(-d)}}{8}. \tag{129}$$

Using (128) and (129) in (120), and combining the result with (119), (86), and (16) completes the proof.

Lemma 6: For any $r>0$

$$0\le\tilde{f}_i^-(r,u)\le\tilde{g}_i^-(u),\quad u\in\mathbb{R} \tag{130}$$

where

$$\tilde{g}_i^-(u)\triangleq\frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{131}$$

and $\hat{d}$ is the maximum ED of the constellation. Furthermore,

$$\int_{-\infty}^{\infty}\tilde{g}_i^-(u)\,\mathrm{d}u=\frac{\hat{d}^2(M-1)^2}{p_i^2}<\infty. \tag{132}$$

Proof: The first inequality in (130) follows directly from (124). To prove the second inequality in (130), we use $\mathrm{e}^{-\frac{u^2}{r^2}}\le 1$, $h(r^2-u)\le 1$, and $\delta\le\hat{d}$ to upper-bound (124) as

$$\tilde{f}_i^-(r,u)\le\hat{d}^2\mathrm{e}^{2u}\left(\frac{\sum_{\delta\in\mathcal{D}_i^*}R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}{1+\sum_{\delta\in\mathcal{D}_i^*}R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}\right)^2 \tag{133}$$

$$=\hat{d}^2\mathrm{e}^{2u}\left(1+\frac{1}{\sum_{\delta\in\mathcal{D}_i^*}R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}\right)^{-2}. \tag{134}$$

Since $R_{P_X}^{(i)}(\delta)<1/p_i$ and $\mathrm{e}^{-4r^2U(\delta)}\le 1$, we can further upper-bound (134) as

$$\tilde{f}_i^-(r,u)<\hat{d}^2\mathrm{e}^{2u}\left(1+\frac{p_i}{\sum_{\delta\in\mathcal{D}_i^*}\mathrm{e}^{-4u\frac{\delta}{\hat{d}}}}\right)^{-2} \tag{135}$$

$$<\frac{\hat{d}^2\mathrm{e}^{2u}}{p_i^2}\left(1+\frac{1}{\sum_{\delta\in\mathcal{D}_i^*}\mathrm{e}^{-4u\frac{\delta}{\hat{d}}}}\right)^{-2} \tag{136}$$

where to pass from (135) to (136) we used $p_i<1$.

If $u\ge 0$, we have

$$\tilde{f}_i^-(r,u)<\frac{\hat{d}^2\mathrm{e}^{2u}}{p_i^2}\left(1+\frac{1}{(M-1)\mathrm{e}^{-4u}}\right)^{-2} \tag{137}$$

$$<\frac{\hat{d}^2\mathrm{e}^{2u}}{p_i^2}\left(\frac{1}{(M-1)\mathrm{e}^{-4u}}\right)^{-2} \tag{138}$$

$$=\frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-6u} \tag{139}$$

$$<\frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{140}$$

where to pass from (136) to (137) all the exponentials are replaced by the one with the largest argument.

If $u\le 0$, (136) can be upper-bounded as

$$\tilde{f}_i^-(r,u)<\frac{\hat{d}^2}{p_i^2}\mathrm{e}^{2u} \tag{141}$$

$$\le\frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{142}$$

where (141) follows from discarding the sum of exponentials in (136). Combining (140) and (142) proves (131). □

Appendix C
Proof of Theorem 3

Using Bayes' rule, $\hat{X}^{\mathrm{MAP}}(y)$ in (9) can be expressed as

$$\hat{X}^{\mathrm{MAP}}(y)=\operatorname*{argmax}_{x\in\mathcal{X}}\left\{f_{Y|X}(y|x)P_X(x)\right\} \tag{143}$$

$$=x_j,\quad \text{if } y\in\mathcal{Y}_j(\rho), \tag{144}$$

where $\mathcal{Y}_j(\rho)$ is the decision region for the symbol x_j with $j=1,\dots,M$. For sufficiently large ρ, these decision regions can be written as

$$\mathcal{Y}_j(\rho)\triangleq\{y\in\mathbb{R}:\beta_{j-1}(\rho)\le y<\beta_j(\rho)\} \tag{145}$$

where $\beta_l(\rho)$ with $l=0,\dots,M$ are the $M+1$ thresholds defining the M regions, i.e.,

$$\beta_l(\rho)=\begin{cases}-\infty, & l=0\\ \frac{\log(p_l/p_{l+1})}{\sqrt{\rho}(x_{l+1}-x_l)}+\frac{\sqrt{\rho}(x_{l+1}+x_l)}{2}, & l=1,\dots,M-1\\ +\infty, & l=M\end{cases} \tag{146}$$

where $\beta_l(\rho)$ for $l=1,\dots,M-1$ in (146) is obtained by using (143) and by solving

$$p_l f_{Y|X}(\beta_l(\rho)|x_l)=p_{l+1}f_{Y|X}(\beta_l(\rho)|x_{l+1}). \tag{147}$$

First we introduce a lemma with general asymptotic results on the thresholds given in (146). This Lemma will be used in this proof as well as in the proof of Theorem 6 (Sec. IV).

Lemma 7: For any P_X and $i\in\mathcal{I}_{\mathcal{X}}$

$$\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(|\beta_l(\rho)-\sqrt{\rho}x_i|\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}=\begin{cases}\sqrt{R_{P_X}^{(i)}(d)}, & \text{if } l=i-1\\ \sqrt{R_{P_X}^{(i)}(-d)}, & \text{if } l=i\\ 0, & \text{if } l\notin\{i-1,i\}\end{cases} \tag{148}$$

where $\beta_l(\rho)$ is given by (146) and $R_{P_X}^{(i)}(\delta)$ by (87).

Proof: We use (146) to obtain

$$\beta_l(\rho) - \sqrt{\rho}x_i = \frac{\log(p_l/p_{l+1})}{\sqrt{\rho}(x_{l+1}-x_l)} + \frac{\sqrt{\rho}\epsilon_{i,l}}{2} \tag{149}$$

where for any i, l

$$\epsilon_{i,l} \triangleq x_{l+1} + x_l - 2x_i. \tag{150}$$

Using (149) and (79), we form the ratio

$$\frac{G\left(|\beta_l(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)} = \frac{\rho d(x_{l+1}-x_l)}{|2\log(p_l/p_{l+1}) + \rho\epsilon_{i,l}(x_{l+1}-x_l)|} \cdot \exp\left(-\frac{\left(\log\frac{p_l}{p_{l+1}}\right)^2}{2\rho(x_{l+1}-x_l)^2} - \frac{\epsilon_{i,l}\log\frac{p_l}{p_{l+1}}}{2(x_{l+1}-x_l)} - \frac{\rho(\epsilon_{i,l}^2 - d^2)}{8}\right). \tag{151}$$

It follows from (150) that $|\epsilon_{i,l}| \geq x_{l+1} - x_l \geq d$ for all i, l, which implies that the limit

$$\lim_{\rho\to\infty} \frac{G\left(|\beta_l(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)} \tag{152}$$

exists. We distinguish between three cases:

(i) If $i = l$ and $x_{l+1} - x_l = d$, then $\epsilon_{i,l} = x_{l+1} - x_l = d$ and the limit in (152) is $\mathrm{e}^{-\log(p_l/p_{l+1})/2} = \sqrt{p_{l+1}/p_l}$.
(ii) If $i = l+1$ and $x_{l+1} - x_l = d$, then $\epsilon_{i,l} = x_l - x_{l+1} = -d$ and the limit in (152) is $\sqrt{p_l/p_{l+1}}$.
(iii) In all other cases, $|\epsilon_{i,l}| > d$ and the limit in (152) is zero.

Combining the three cases and slightly changing notation yields

$$\lim_{\rho\to\infty} \frac{G\left(|\beta_l(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)} = \begin{cases} \sqrt{\frac{p_{i+1}}{p_i}}, & \text{if } l = i \text{ and } x_{l+1} - x_l = d \\ \sqrt{\frac{p_{i-1}}{p_i}}, & \text{if } l = i-1 \text{ and } x_{l+1} - x_l = d \\ 0, & \text{otherwise} \end{cases}. \tag{153}$$

Finally, applying (87) and (80) completes the proof. □

Returning to the proof of Theorem 3, using (144) and (145), the SEP in (8) is expressed as

$$\mathsf{S}_{P_X}(\rho) = \sum_{i\in\mathcal{I}_\mathcal{X}} p_i \Pr\{Y \notin \mathcal{Y}_i(\rho) | X = x_i\} \tag{154}$$

$$= \sum_{i\in\mathcal{I}_\mathcal{X}} p_i \big(\mathsf{Q}\left(\beta_i(\rho) - \sqrt{\rho}x_i\right) + \mathsf{Q}\left(\sqrt{\rho}x_i - \beta_{i-1}(\rho)\right)\big) \tag{155}$$

which gives

$$\lim_{\rho\to\infty} \frac{\mathsf{S}_{P_X}(\rho)}{\mathsf{Q}(\sqrt{\rho}d/2)} = \sum_{i\in\mathcal{I}_\mathcal{X}} p_i \left(\lim_{\rho\to\infty} \frac{\mathsf{Q}\left(\beta_i(\rho) - \sqrt{\rho}x_i\right)}{\mathsf{Q}(\sqrt{\rho}d/2)} + \lim_{\rho\to\infty} \frac{\mathsf{Q}\left(\sqrt{\rho}x_i - \beta_{i-1}(\rho)\right)}{\mathsf{Q}(\sqrt{\rho}d/2)}\right) \tag{156}$$

$$= \sum_{i\in\mathcal{I}_\mathcal{X}} p_i \left(\sqrt{R_{P_X}^{(i)}(-d)} + \sqrt{R_{P_X}^{(i)}(d)}\right) \tag{157}$$

where to pass from (156) to (157) we used Lemma 7 twice, observing that the arguments of both Q-functions are positive for large enough ρ. The proof of Theorem 3 is completed by using (86) in (157) together with (16).

Appendix D
Proof of Theorem 4

The following lemma will be used in this proof as well as in the proof of Theorem 5.

Lemma 8:

$$\sum_{k=1}^{m} \left(B_{P_X} - \sum_{b\in\mathcal{B}} P_{Q_k}(b) B_{P_{X_{k,b}}}\right) = D_{P_X,\boldsymbol{l}} \tag{158}$$

where $D_{P_X,\boldsymbol{l}}$ is given by (41), B_{P_X} is given by (16),

$$B_{P_{X_{k,b}}} = \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w\in\mathcal{W}} B_{P_{X_{k,b}}}^{(i)}(wd) \tag{159}$$

and

$$B_{P_{X_{k,b}}}^{(i)}(\delta) = \begin{cases} \sqrt{P_{X_{k,b}}(x_j) P_{X_{k,b}}(x_i)}, & \text{if } \exists x_j \in \mathcal{X}_{k,b} : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases} \tag{160}$$

with $P_{X_{k,b}}(x)$ given by (34).

Proof: From (34), $p_i = P_{Q_k}(b) P_{X_{k,b}}(x_i)$ for any $b \in \mathcal{B}$, $k = 1, \ldots, m$ and $i \in \mathcal{I}_{\mathcal{X}_{k,b}}$. Hence, using (160) and (42) gives

$$D_{P_{X_{k,b}}}^{(i)}(\delta) = P_{Q_k}(b) B_{P_{X_{k,b}}}^{(i)}(\delta). \tag{161}$$

Using (159) and (161) together with (16),

$$\sum_{k=1}^{m} \left(B_{P_X} - \sum_{b\in\mathcal{B}} P_{Q_k}(b) B_{P_{X_{k,b}}}\right) = \sum_{k=1}^{m} \sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w\in\mathcal{W}} \left(B_{P_X}^{(i)}(wd) - D_{P_{X_{k,b}}}^{(i)}(wd)\right) \tag{162}$$

$$= \sum_{k=1}^{m} \sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} \sum_{w\in\mathcal{W}} D_{P_{X_{k,\overline{b}}}}^{(i)}(wd) \tag{163}$$

where (163) is obtained using (43). The proof is completed by comparing (163) with (41). □

Using the expression for the BICM-GMI (45), we have

$$H_{P_X} - \mathsf{I}_{P_X,\boldsymbol{l}}^{\mathrm{BI}}(\rho) = \sum_{k=1}^{m} (H_{P_X} - \mathsf{I}_{P_X}(\rho)) - \sum_{k=1}^{m} \sum_{b\in\mathcal{B}} P_{Q_k}(b)(H_{P_{X_{k,b}}} - \mathsf{I}_{P_{X_{k,b}}}(\rho)) - (m-1)H_{P_X} + \sum_{k=1}^{m} \sum_{b\in\mathcal{B}} P_{Q_k}(b) H_{P_{X_{k,b}}}. \tag{164}$$

The last term on the r.h.s. of (164) is zero because

$$\begin{aligned}
&\sum_{k=1}^{m}\sum_{b\in\mathcal{B}} P_{Q_k}(b) H_{P_{X_{k,b}}} \\
&= -\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} P_{Q_k}(b) P_{X_{k,b}}(x_i)\log P_{X_{k,b}}(x_i) && (165)\\
&= -\sum_{k=1}^{m}\sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \log\frac{p_i}{P_{Q_k}(q_{i,k})} && (166)\\
&= mH_{P_X} + \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \sum_{k=1}^{m}\log P_{Q_k}(q_{i,k}) && (167)\\
&= mH_{P_X} + \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \log\prod_{k=1}^{m} P_{Q_k}(q_{i,k}) && (168)\\
&= mH_{P_X} - H_{P_X} && (169)
\end{aligned}$$

where to pass from (165) to (166) we used (34), and to pass from (168) to (169) we used (33).

We divide both sides of (164) by $\mathsf{Q}(\sqrt{\rho}d/2)$ and take the limit as $\rho\to\infty$. For the first two terms, we change the order of summation and limit and apply Theorem 1 to each term. This gives

$$\begin{aligned}
&\lim_{\rho\to\infty}\frac{H_{P_X} - \mathsf{I}^{\mathrm{BI}}_{P_X,\boldsymbol{\iota}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}\\
&\qquad = \pi\sum_{k=1}^{m}\left(B_{P_X} - \sum_{b\in\mathcal{B}} P_{Q_k}(b) B_{P_{X_{k,b}}}\right) \qquad (170)
\end{aligned}$$

which in view of Lemma 8 completes the proof.

Appendix E
Proof of Theorem 6

The BEP in (53) is expressed as

$$\begin{aligned}
&\mathsf{B}_{P_X,\boldsymbol{\iota}}(\rho)\\
&= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i \Pr\{\hat{Q}_k^{\mathrm{MAP}}(Y)\neq q_{k,i}|X=x_i\} && (171)\\
&= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i \Pr\left\{Y\in\bigcup_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}}\mathcal{Y}_j(\rho)\,\middle|\,X=x_i\right\} && (172)
\end{aligned}$$

where (171) follows from applying the law of total probability, and (172) follows from considering all the decision regions that include a constellation point labeled by $\overline{b}$ at bit position k, by using the fact that $\mathcal{Y}_j(\rho)$ are disjoint, and $\cup_{j\in\mathcal{I}_{\mathcal{X}}}\mathcal{Y}_j = \mathbb{R}$. Furthermore, we express (172) in terms of pairwise error probabilities (PEP) as

$$\begin{aligned}
&\mathsf{B}_{P_X,\boldsymbol{\iota}}(\rho)\\
&= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i \sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}} \Pr\{Y\in\mathcal{Y}_j(\rho)|X=x_i\} && (173)\\
&= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i \sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}} \mathrm{PEP}_{i,j}(\rho) && (174)
\end{aligned}$$

where

$$\mathrm{PEP}_{i,j}(\rho) \triangleq \mathsf{Q}\left(\beta_{j-1}(\rho) - \sqrt{\rho}x_i\right) - \mathsf{Q}\left(\beta_j(\rho) - \sqrt{\rho}x_i\right). \qquad (175)$$

Since for any $i\in\mathcal{I}_{\mathcal{X}_{k,b}}$ the innermost sum in (174) considers only $j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}$ (i.e., $j\neq i$), we express it as

$$\begin{aligned}
\mathsf{B}_{P_X,\boldsymbol{\iota}}(\rho) = \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i\Bigg(&\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j<i}\mathrm{PEP}_{i,j}(\rho)\\
&+\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j>i}\mathrm{PEP}_{i,j}(\rho)\Bigg). \qquad (176)
\end{aligned}$$

For $j<i$ and sufficiently large ρ, the arguments of the Q-functions in (175) are negative. Thus, we use $\mathsf{Q}(-x) = 1-\mathsf{Q}(x)$ to express $\mathrm{PEP}_{i,j}(\rho)$ for $j<i$ as

$$\mathrm{PEP}_{i,j}(\rho) = \mathsf{Q}\left(\sqrt{\rho}x_i - \beta_j(\rho)\right) - \mathsf{Q}\left(\sqrt{\rho}x_i - \beta_{j-1}(\rho)\right). \qquad (177)$$

By using (175) and (177) in (176), dividing both sides of (176) by $\mathsf{Q}\left(\sqrt{\rho}d/2\right)$, and taking the limit as $\rho\to\infty$, we obtain

$$\lim_{\rho\to\infty}\frac{\mathsf{B}_{P_X,\boldsymbol{\iota}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)} = \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i\sum_{l=1}^{4} S_l \qquad (178)$$

where

$$S_1 = \sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j<i}\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(\sqrt{\rho}x_i - \beta_j(\rho)\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}, \qquad (179)$$

$$S_2 = -\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j<i}\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(\sqrt{\rho}x_i - \beta_{j-1}(\rho)\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}, \qquad (180)$$

$$S_3 = \sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j>i}\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(\beta_{j-1}(\rho) - \sqrt{\rho}x_i\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}, \qquad (181)$$

$$S_4 = -\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}},j>i}\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(\beta_j(\rho) - \sqrt{\rho}x_i\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}. \qquad (182)$$

where all the arguments of the Q-functions in (179)–(182) are positive for large enough ρ.

Due to Lemma 7, we conclude that $S_4 = 0$ and that S_3 could be nonzero only due to the contribution of the term $j = i+1$. To compute S_2, we use $\mathsf{Q}\left(\sqrt{\rho}x_i - \beta_{j-1}(\rho)\right) = \mathsf{Q}\left(|\beta_{j-1}(\rho) - \sqrt{\rho}x_i|\right)$ and Lemma 7 to obtain $S_2 = 0$. Similarly, using Lemma 7, we we conclude that the only nonzero contribution to S_1 can can come from the term

$j = i - 1$. Combining these results and using the counting function (11), we express (178) as

$$\lim_{\rho\to\infty} \frac{\mathsf{B}_{P_X,\boldsymbol{\iota}}(\rho)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i\left(A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(-d)\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(\beta_i(\rho)-\sqrt{\rho}x_i\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}\right.$$

$$\left. + A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(d)\lim_{\rho\to\infty}\frac{\mathsf{Q}\left(|\beta_{i-1}(\rho)-\sqrt{\rho}x_i|\right)}{\mathsf{Q}\left(\sqrt{\rho}d/2\right)}\right) \quad (183)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} p_i\left(A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(-d)\sqrt{R^{(i)}_{P_X}(-d)}\right.$$

$$\left. + A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(d)\sqrt{R^{(i)}_{P_X}(d)}\right) \quad (184)$$

where to pass from (183) to (184) we used Lemma 7. Furthermore, by combining (87) and (42) we obtain

$$p_i A^{(i)}_{\mathcal{X}_{k,\overline{b}}}(\delta)\sqrt{R^{(i)}_{P_X}(\delta)} = D^{(i)}_{P_{X_{k,\overline{b}}}}(\delta) \quad (185)$$

which combined with (184) and (41) completes the proof.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communications," *Bell System Technical Journal*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, June 1990.

[3] ——, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[4] V. V. Prelov and S. Verdú, "Second-order asymptotics of mutual information," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1567–1580, Aug. 2004.

[5] A. Lozano, A. M. Tulino, and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[6] F. Pérez-Cruz, M. R. D. Rodrigues, and S. Verdú, "MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1070–1084, Mar. 2010.

[7] Y. Wu and S. Verdú, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug. 2011.

[8] D. Duyck, J. J. Boutros, and M. Moeneclaey, "Precoding for outage probability minimization on block fading channels," *submitted to IEEE Trans. Inf. Theory*, Oct. 2012, available at http://arxiv.org/abs/1103.5348.

[9] M. R. D. Rodrigues, "On the constrained capacity of multi-antenna fading coherent channels with discrete inputs," in *IEEE International Symposium on Information Theory (ISIT)*, Saint Petersburg, Russia, July-Aug. 2011.

[10] ——, "Characterization of the constrained capacity of multiple-antenna fading coherent channels driven by arbitrary inputs," in *IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, July 2012.

[11] ——, "Multiple-antenna fading coherent channels with arbitrary inputs: Characterization and optimization of the reliable information transmission rate," Oct. 2012, available at http://arxiv.org/abs/1210.6777.

[12] A. G. C. P. Ramos and M. R. D. Rodrigues, "Coherent fading channels driven by arbitrary inputs: Asymptotic characterization of the constrained capacity and related information- and estimation-theoretic quantities," Oct. 2012, available at http://arxiv.org/abs/1210.4505.

[13] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.

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

[15] 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.

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

[17] C. Stierstorfer and R. F. H. Fischer, "(Gray) Mappings for bit-interleaved coded modulation," in *IEEE Vehicular Technology Conference (VTC-Spring)*, Dublin, Ireland, Apr. 2007.

[18] C. Stierstorfer, "A bit-level-based approach to coded multicarrier transmission," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, 2009, available at http://www.opus.ub.uni-erlangen.de/opus/volltexte/2009/1395/.

[19] A. Martinez, A. Guillén i Fàbregas, and G. Caire, "Bit-interleaved coded modulation in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5447–5455, Dec. 2008.

[20] C. Stierstorfer and R. F. H. Fischer, "Asymptotically optimal mappings for BICM with M-PAM and M^2-QAM," *IET Electronics Letters*, vol. 45, no. 3, pp. 173–174, Jan. 2009.

[21] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, Oct. 2011.

[22] ——, "Signal shaping for BICM at low SNR," *IEEE Trans. Inf. Theory*, 2013, to appear.

[23] A. Alvarado, F. Brännström, and E. Agrell, "High SNR bounds for the BICM capacity," in *IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011.

[24] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[25] D. Guo, "Gaussian channels: Information, estimation and multiuser detection," Ph.D. dissertation, Princeton University, Princeton, New Jersey, Nov. 2004.

[26] J. B. Anderson and A. Svensson, *Coded Modulation Systems*. Springer, 2003.

[27] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar. 1953.

[28] 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.

[29] 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.

[30] A. Guillén i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *IEEE Information Theory Workshop (ITW)*, Dublin, Ireland, Aug.–Sep. 2010.

[31] ——, "Derivative of BICM mutual information," *IET Electronics Letters*, vol. 43, no. 22, pp. 1219–1220, Oct. 2007.

[32] M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, "On the exact BER of bit-wise demodulators for one-dimensional constellations," *IEEE Trans. Commun.*, 2013 (to appear), available at http://arxiv.org/abs/1206.2478.

[33] ——, "General BER expression for one-dimensional constellations," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012, available at http://arxiv.org/abs/1210.8326.

[34] ETSI, "Digital video broadcasting (DVB); framing structure, channel coding and modulation for digital terrestrial television," ETSI, Tech. Rep. ETSI EN 300 744 V1.6.1 (2009-01), Jan. 2009.

[35] U. Madhow, *Fundamentals of Digital Communication*, 1st ed. Cambridge University Press, 2008.

[36] A. Lapidoth, *A Foundation in Digital Communication*, 1st ed. Cambridge University Press, 2009.

[37] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, 1987.

[38] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 6th ed. New York, NY: Academic Press, 2000.