Alvarado, Alex; Brännström, Fredrik; Agrell, Erik; Koch, Tobias. High-SNR Asymptotics of Mutual Information for Discrete Constellations with Applications to BICM, in: *IEEE Transactions on Information Theory, 60(2), pp. 1061-1076, Feb. 2014*

DOI: https://doi.org/10.1109/TIT.2013.2291865

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

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

***Abstract*—Asymptotic expressions of the mutual information between any discrete input and the corresponding output of the scalar additive white Gaussian noise channel are presented in the limit as the signal-to-noise ratio (SNR) tends to infinity. Asymptotic expressions of the symbol-error probability (SEP) and the minimum mean-square error (MMSE) achieved by estimating the channel input given the channel output are also developed. It is shown that for any input distribution, the conditional entropy of the channel input given the output, MMSE and SEP have an asymptotic behavior proportional to the Gaussian Q-function. The argument of the Q-function depends only on the minimum Euclidean distance (MED) of the constellation and the SNR, and the proportionality constants are functions of the MED and the probabilities of the pairs of constellation points at MED. The developed expressions are then generalized to study the high-SNR behavior of the generalized mutual information (GMI) for bit-interleaved coded modulation (BICM). By means of these asymptotic expressions, the long-standing conjecture that Gray codes are the binary labelings that maximize the BICM-GMI at high SNR is proven. It is further shown that for any equally spaced constellation whose size is a power of two, there always exists an anti-Gray code giving 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

We consider the discrete-time, real-valued, additive white Gaussian noise (AWGN) channel

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

where X is the transmitted symbol; Z is a Gaussian random variable, independent of X, with zero mean and unit variance; and $\rho > 0$ is an arbitrary scale factor. The capacity of the AWGN channel (1) under an average-power constraint $\rho\mathbb{E}_X[X^2] \leq \gamma$ is given by [1]

$$C^{\text{awgn}}(\gamma) = \frac{1}{2}\log(1+\gamma) \tag{2}$$

where γ can be viewed as the maximal allowed signal-to-noise ratio (SNR). 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 distribution of X 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 estimating X from Y. In [8, App. 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, asymptotic expansions for the MMSE and MI for scalar and vectorial coherent fading channels were recently derived in [12].

In this paper, we study the 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 conditional entropy of X given Y, the MMSE, and the SEP have an asymptotic behavior proportional to $Q\left(\sqrt{\rho}d/2\right)$, where $Q(\cdot)$ 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

Research supported by the European Community's Seventh's Framework Programme (FP7/2007-2013) under grant agreements No. 271986 and No. 333680, by the Swedish Research Council, Sweden (under grants #621-2006-4872 and #621-2011-5950) and by the Ministerio de Economía y Competitividad of Spain (TEC2009-14504-C02-01, CSD2008-00010, and TEC2012-38800-C03-01). This work was presented in part at the IEEE International Symposium on Information Theory, Istanbul, Turkey, July 2013, and at the IEEE Communication Theory Workshop, Phuket, Thailand, June 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).

SNR. Furthermore, in contrast to previous works, we provide closed-form expressions for the coefficients in front of the Q-functions, thereby characterizing the asymptotic behavior of the MI, MMSE, and SEP more accurately.

While these asymptotic 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 in [16] as a mismatched decoder. 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 the sense that it maximizes 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 indeed 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 converge to their asymptotes proportionally to $Q\left(\sqrt{\rho}d/2\right)$.

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 $\mathcal{X}$. An exception is the set of real numbers, which is denoted by $\mathbb{R}$. 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 (1). The transmitted symbols X belong to a real-valued, one-dimensional *constellation* $\mathcal{X} \triangleq \{x_1, x_2, \ldots, x_M\}$ where $M = 2^m$, and they are, without loss of generality, assumed to be distinct and ordered, i.e., $x_1 < x_2 < \cdots < x_M$. Each symbol is transmitted with probability $p_i \triangleq P_X(x_i)$, $0 < p_i < 1$, and the vector of probabilities $\boldsymbol{p} \triangleq [p_1, \ldots, p_M]$ is called the *input distribution*. We assume that neither the constellation nor the input distribution depends on ρ. We denote the set of indices in $\mathcal{X}$ and $\boldsymbol{p}$ with $\mathcal{I}_\mathcal{X} \triangleq \{1, \ldots, M\}$.

The transmitted average symbol energy is

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

It follows that the SNR 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

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

The entropy of the random variable X is defined as

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

the MI between X and Y, $I(X;Y)$, as

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

and the MMSE as

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

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

$$S_{P_X}(\rho) \triangleq \Pr\{\hat{X}^\mathrm{map}(Y) \neq X\} \tag{8}$$

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.

The MED of the constellation is defined as

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

We further define $A_\mathcal{X}$ as twice the number of pairs of constellation points at MED, i.e.,

$$A_\mathcal{X} \triangleq \sum_{i \in \mathcal{I}_\mathcal{X}} \sum_{\substack{j \in \mathcal{I}_\mathcal{X} \\ |x_i - x_j| = d}} 1. \tag{11}$$

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

$$2 \leq A_\mathcal{X} \leq 2(M - 1). \tag{12}$$

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

$$A_\mathcal{E} = 2(M - 1). \tag{13}$$

Finally, for a given P_X, we define the constant

$$B_{P_X} \triangleq \sum_{i\in\mathcal{I}_\mathcal{X}} \sum_{\substack{j\in\mathcal{I}_\mathcal{X} \\ |x_i-x_j|=d}} \sqrt{p_j p_i}. \tag{14}$$

For a uniform input distribution, $P_X = P_X^{\mathrm{u}}$, and thus,

$$B_{P_X^{\mathrm{u}}} = \frac{A_\mathcal{X}}{M}. \tag{15}$$

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 (11) is $A_\mathcal{X} = 4$ (two pairs of constellation points at MED), and B_{P_X} in (14) is $B_{P_X} = 2\sqrt{p_1 p_2} + 2\sqrt{p_3 p_4} \approx 0.98$.

III. High-SNR Asymptotics

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

$$\frac{\mathrm{d}}{\mathrm{d}\rho} I_{P_X}(\rho) = \frac{1}{2} M_{P_X}(\rho). \tag{16}$$

Exploiting this MI-MMSE relation, bounds on the MI can be used to derive bounds on the MMSE and *vice versa*. The relationship in (16) will be used in the proof of Theorem 1.

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. We further present exact asymptotic expressions for the SEP.

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 $H_{P_X} - I_{P_X}(\rho)$, i.e., by analyzing the conditional entropy of X given Y. Theorem 1 is the main result of this paper and characterizes the high-SNR behavior of the conditional entropy $H(X|Y) = H_{P_X} - I_{P_X}(\rho)$.

Theorem 1: For any P_X

$$\lim_{\rho\to\infty} \frac{H_{P_X} - I_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \pi B_{P_X} \tag{17}$$

where B_{P_X} is given by (14).

Proof: See Appendix A. □

Similar to Theorem 1, we have the following asymptotic expressions for the MMSE and the SEP.

Theorem 2: For any P_X

$$\lim_{\rho\to\infty} \frac{M_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} B_{P_X}. \tag{18}$$

Proof: See Appendix B. □

Theorem 3: For any P_X

$$\lim_{\rho\to\infty} \frac{S_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = B_{P_X}. \tag{19}$$

Proof: See Appendix C. □

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

$$I_{P_X}(\rho) \approx H_{P_X} - \pi B_{P_X} Q\left(\sqrt{\rho}d/2\right) \tag{20}$$

$$M_{P_X}(\rho) \approx \frac{\pi d^2}{4} B_{P_X} Q\left(\sqrt{\rho}d/2\right) \tag{21}$$

$$S_{P_X}(\rho) \approx B_{P_X} Q\left(\sqrt{\rho}d/2\right). \tag{22}$$

The results in (20)–(22) show that for any input distribution, the conditional entropy, MMSE, and SEP have the same high-SNR behavior: i.e., they are all proportional to a Gaussian Q-function, where the proportionality constants depend on the input distribution and, in the case of the MMSE, also on the MED of the constellation.

Remark 1: While the results presented in this section were derived only for one-dimensional (real-valued) 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.

Corollary 4: For any $\mathcal{X}$ and a uniform input distribution

$$\lim_{\rho\to\infty} \frac{\log M - I_{P_X^{\mathrm{u}}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \pi \frac{A_\mathcal{X}}{M} \tag{23}$$

$$\lim_{\rho\to\infty} \frac{M_{P_X^{\mathrm{u}}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} \frac{A_\mathcal{X}}{M} \tag{24}$$

$$\lim_{\rho\to\infty} \frac{S_{P_X^{\mathrm{u}}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{A_\mathcal{X}}{M} \tag{25}$$

where $A_\mathcal{X}$ is given by (11).

Proof: From Theorems 1–3 and (15). □

The expression (25) corresponds to the well-known high-SNR approximation for the SEP [26, eq. (2.3-29)]. Corollary 4 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.

For MPAM and a uniform input distribution ($P_X = P_X^{\mathrm{eu}}$), it follows from Corollary 4 and (13) that

$$I_{P_X^{\mathrm{eu}}}(\rho) \approx \log M - \frac{2\pi(M-1)}{M} Q\left(\sqrt{\rho}d/2\right) \tag{26}$$

$$M_{P_X^{\mathrm{eu}}}(\rho) \approx \frac{6\pi E_{\mathrm{s}}}{M(M+1)} Q\left(\sqrt{\rho}d/2\right) \tag{27}$$

$$S_{P_X^{\mathrm{eu}}}(\rho) \approx \frac{2(M-1)}{M} Q\left(\sqrt{\rho}d/2\right). \tag{28}$$

Table I summarizes the results obtained in Theorems 1–3, Corollary 4, and (26)–(28).

Example 2: In Fig. 1, we show $\log M - I_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM with uniform input distributions[1] together with

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

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

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

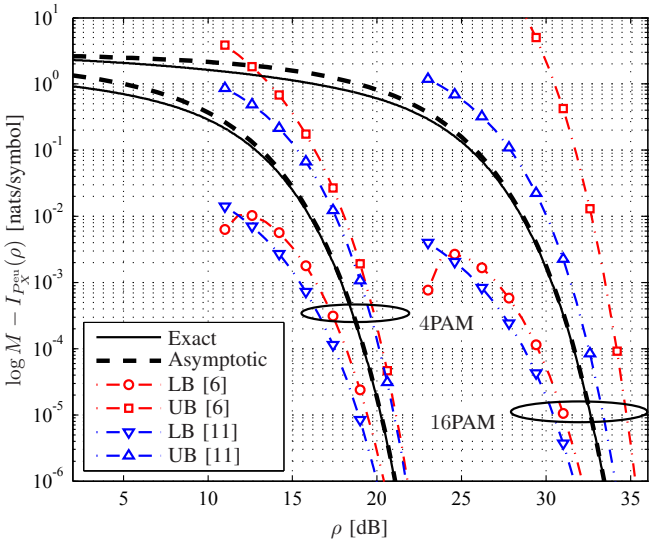

Fig. 1. $\log M - I_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_{\mathrm{s}} = 1$) and the asymptotic expression in (26) (thick dashed lines). The lower and upper bounds [6, eq. (34)–(35)] and [11, eq. (17)–(19)] are also shown.

the asymptotic expression in (26). We also show the lower and upper bounds derived in [6, eq. (34)–(35)] and [11, eq. (17)–(19)]. Observe that (26) approximates $I_{P_X^{\mathrm{eu}}}(\rho)$ 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. Again, the asymptotic expression (27) approximates the MMSE accurately for a large range of SNR. For other examples with unequally spaced 4-ary constellations and nonuniform input distributions see [27, Example 3].

Remark 2: It follows from Corollary 4 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 $A_{\mathcal{X}}$. It is easy to see that the one-dimensional constellation that maximizes the MED for a given E_{s} is the MPAM constellation ($\mathcal{X} = \mathcal{E}$).

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

BICM can be viewed as a pragmatic approach for coded modulation. In BICM (see Fig. 3), the encoder is realized as a serial concatenation of a binary encoder, a bit-level interleaver,

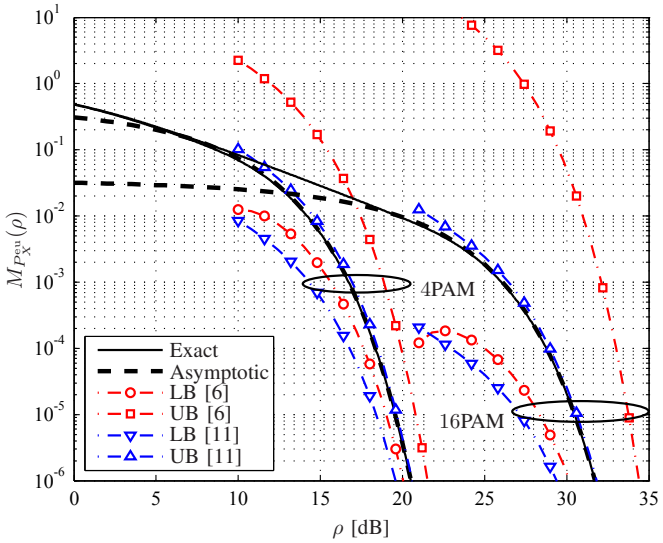

Fig. 2. $M_{P_X^{\mathrm{eu}}}(\rho)$ for 4PAM and 16PAM (solid lines) constellations (normalized to $E_{\mathrm{s}} = 1$) and the asymptotic expression in (27) (thick dashed lines). The lower and upper bounds [6, eq. (30)–(31)] and [11, eq. (13)–(15)] are also shown.

and a memoryless mapper; see [13]–[15] for more details. A key element in BICM is the memoryless mapper $\Phi : \{0,1\}^m \to \mathcal{X}$, which maps the coded bits $\boldsymbol{Q} = [Q_1, \ldots, Q_m]$ to constellation symbols. At the receiver side, the demapper computes a bit metric, typically in the form of a logarithmic likelihood ratio (LLR)

$$\Lambda_k \triangleq \log \frac{f_{Y|Q_k}(y|1)}{f_{Y|Q_k}(y|0)} \tag{29}$$

for $k = 1, \ldots, m$. The vector of LLRs $\boldsymbol{\Lambda} = [\Lambda_1, \ldots, \Lambda_m]$ is deinterleaved and then decoded.

The BICM-GMI is an achievable rate for BICM, and thus, is an important quantity for such systems. In this section, we generalize the results in Sec. III to obtain asymptotic expressions for the BICM-GMI. We further study the relationship between the BICM-GMI and the BEP as well as the derivative of the BICM-GMI with respect to ρ. Finally, we show that at high SNR, GCs maximize BICM-GMI for one-dimensional constellations and 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 $\boldsymbol{q}_i = [q_{i,1}, \ldots, q_{i,m}] \in \{0,1\}^m$ 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 $b \in \{0,1\}$, 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 6 subconstellations for an 8PAM constellation labeled by the NBC $\boldsymbol{l} = [0, 1, 2, 3, 4, 5, 6, 7]$ [21, Sec. II-B], as well as the corresponding values of $\mathcal{I}_{\mathcal{X}_{k,b}}$ and $A_{\mathcal{X}_{k,b}}$ (defined below).

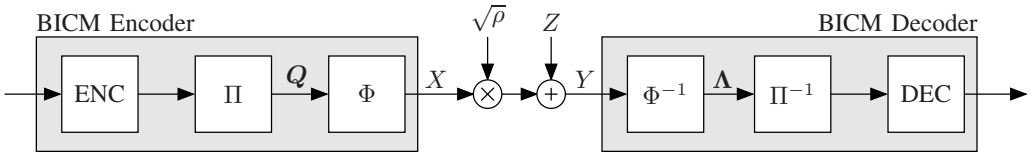

Fig. 3. A BICM scheme: The BICM encoder is formed by a serial concatenation of a binary encoder (ENC), a bit-level interleaver (Π), and a memoryless mapper (Φ). The BICM decoder is based on a demapper (Φ^{-1}) that computes logarithmic likelihood ratios, a deinterleaver (Π^{-1}), and a channel decoder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

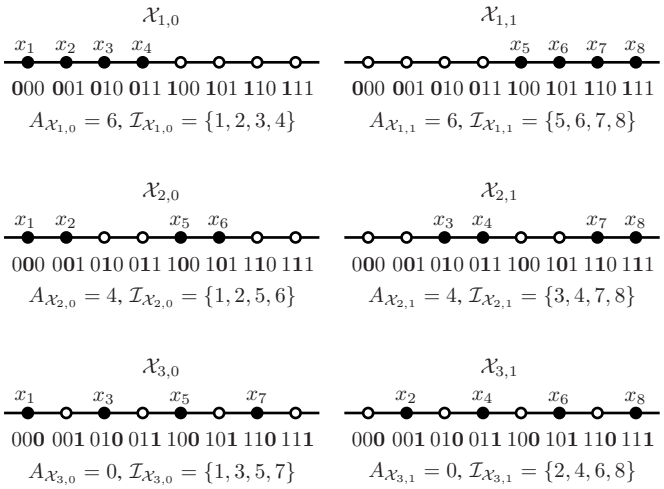

Fig. 4. Subconstellations $\mathcal{X}_{k,b}$ (black circles) for 8PAM labeled by the NBC $\boldsymbol{l} = [0,1,2,3,4,5,6,7]$, where the values of $q_{i,k}$ for $k = 1,2,3$ are highlighted in boldface, and where $A_{\mathcal{X}} = 14$ and $C_{\mathcal{X},\boldsymbol{l}} = 22$. The values of $A_{\mathcal{X}_{k,b}}$ and $\mathcal{I}_{\mathcal{X}_{k,b}}$ are also shown.

In BICM, the coded bits $\boldsymbol{Q} = [Q_1, Q_2, \ldots, Q_m]$ at the input of the mapper Φ (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 [28, eq. (8)], [29, eq. (9)]

$$P_X(x_i) = p_i = \prod_{k=1}^{m} P_{Q_k}(q_{i,k}). \tag{30}$$

Using (30), we obtain the conditional probabilities

$$P_{X|Q_k}(x|b) = \begin{cases} \frac{P_X(x)}{P_{Q_k}(b)}, & \text{if } x \in \mathcal{X}_{k,b} \\ 0, & \text{if } x \notin \mathcal{X}_{k,b} \end{cases} \tag{31}$$

for $k = 1, \ldots, m$ and $b \in \{0,1\}$. According to (31), 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 $\mathcal{X}_{k,b}$. We shall use $X_{k,b}$ to denote a random variable with support $\mathcal{X}_{k,b}$ and PMF $P_{X|Q_k}(x|b)$. To shorten notation, we denote this PMF by $P_{X_{k,b}}$.

We next 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 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 (see, for example, $\mathcal{X}_{3,0}$ and $\mathcal{X}_{3,1}$ in Fig. 4). It follows that the bounds (12) on $A_{\mathcal{X}}$ modify to

$$0 \le A_{\mathcal{X}_{k,b}} \le 2\,(M/2 - 1)\,. \tag{32}$$

B. Binary Labelings and Key Quantities for BICM

The NBC [21, Sec. II-B] is defined as the binary labeling $\boldsymbol{l}$ where $l_i = i-1$. It is the only 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 for which $|x_i - x_j| = d$, the binary labels $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ are at Hamming distance one. One of the most popular GCs is the binary reflected Gray code (BRGC) [30]–[32].

To characterize binary labelings we define the constant

$$C_{\mathcal{X},\boldsymbol{l}} \triangleq \sum_{k=1}^{m} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,0}}} \sum_{\substack{j \in \mathcal{I}_{\mathcal{X}_{k,1}} \\ |x_i - x_j| = d}} 2 \tag{33}$$

which corresponds to twice the total number of *different* bits between the labels of constellation symbol pairs at MED. For every given $x_i \in \mathcal{X}_{k,0}$, the inner sum in (33) considers all the constellation points in the subconstellation $\mathcal{X}_{k,1}$ at MED from $x_i \in \mathcal{X}_{k,0}$. According to this interpretation, (33) can alternatively be expressed as

$$C_{\mathcal{X},\boldsymbol{l}} = \sum_{k=1}^{m} \left(A_{\mathcal{X}} - A_{\mathcal{X}_{k,0}} - A_{\mathcal{X}_{k,1}}\right) \tag{34}$$

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, $A_{\mathcal{X}} = 14$ and $C_{\mathcal{X},\boldsymbol{l}} = 22$.

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

$$\begin{aligned} C_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} &= 2\sum_{k=1}^{m}(2^k - 1) \\ &= 2(2M - m - 2) \end{aligned} \tag{35}$$

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 $i = 1, 2, \ldots, M-1$.

Note that, while $A_{\mathcal{X}}$ in (11) depends only on the geometry of the constellation, $C_{\mathcal{X},\boldsymbol{l}}$ in (33) depends on both the geometry of the constellation and the labeling. Since any pair of constellation points at MED will differ in at least one bit, we

have for any $\mathcal{X}$ and $\boldsymbol{l}$

$$C_{\mathcal{X},\boldsymbol{l}} \geq A_{\mathcal{X}}. \tag{36}$$

To state our main results on BICM, and in analogy to (14), we define the constant

$$D_{P_X,\boldsymbol{l}} \triangleq \sum_{k=1}^{m} \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,0}}} \sum_{\substack{j\in\mathcal{I}_{\mathcal{X}_{k,1}} \\ |x_i-x_j|=d}} 2\sqrt{p_j p_i}. \tag{37}$$

Analog to (15), for a uniform input distribution

$$D_{P_X^{\mathrm{u}},\boldsymbol{l}} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{M}. \tag{38}$$

C. Asymptotic Characterization of BICM

The BICM-GMI is an achievable rate for BICM [16] and is one of the key quantities used to analyze BICM systems. For any P_X and $\boldsymbol{l}$, the BICM-GMI is defined as [16, eq. (10)], [21, eqs. (32) and (41)][2]

$$I^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \sum_{k=1}^{m} I(Q_k;Y) \tag{39}$$

$$= \sum_{k=1}^{m}\left(I_{P_X}(\rho) - \sum_{b\in\{0,1\}} P_{Q_k}(b) I_{P_{X_{k,b}}}(\rho)\right). \tag{40}$$

Twice the derivative of $I^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)$ is given by [33, eq. (3)][3]

$$M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho) \triangleq 2\frac{\mathrm{d}I^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)}{\mathrm{d}\rho} \tag{41}$$

$$= \sum_{k=1}^{m}\left(M_{P_X}(\rho) - \sum_{b\in\{0,1\}} P_{Q_k}(b) M_{P_{X_{k,b}}}(\rho)\right). \tag{42}$$

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

$$I_{P_{X_{k,b}}}(\rho) \triangleq \mathbb{E}_{X_{k,b},Y}\left[\log\left(f_{Y|X_{k,b}}(Y|X_{k,b})/f_Y(Y)\right)\right] \tag{43}$$

and

$$M_{P_{X_{k,b}}}(\rho) \triangleq \mathbb{E}_{X_{k,b},Y}[(X_{k,b} - \hat{X}^{\mathrm{me}}_{P_{X_{k,b}}}(Y))^2] \tag{44}$$

$$\hat{X}^{\mathrm{me}}_{P_{X_{k,b}}}(y) \triangleq \mathbb{E}_{X_{k,b}}[X_{k,b}|Y=y] \tag{45}$$

where Y is the random variable resulting from transmitting $X_{k,b}\in\mathcal{X}_{k,b}$ over the AWGN channel (1).

We define the BEP as[4]

$$B_{P_X,\boldsymbol{l}}(\rho) \triangleq \frac{1}{m}\sum_{k=1}^{m} \Pr\{\hat{Q}_k^{\mathrm{map}}(Y) \neq Q_k\} \tag{46}$$

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).[5]

[2]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 (30).

[3]Since the BICM-GMI is not an MI, its derivative is not an MMSE [33]. We thus avoid using the name MMSE, although we do use the MMSE-like notation $M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)$.

[4]Note that (46) is the BEP averaged over the m bit positions, in contrast to the BICM-GMI in (40), which is a sum of m bit-wise MIs.

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

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

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

$$\lim_{\rho\to\infty} \frac{H_{P_X} - I^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \pi D_{P_X,\boldsymbol{l}} \tag{47}$$

where $D_{P_X,\boldsymbol{l}}$ is given by (37).

Proof: See Appendix D. □

Similar to Theorems 2 and 3, we have following asymptotic expressions for $M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)$ and the BEP.

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

$$\lim_{\rho\to\infty} \frac{M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4} D_{P_X,\boldsymbol{l}}. \tag{48}$$

Proof: See Appendix E. □

Theorem 7: For any P_X and $\boldsymbol{l}$

$$\lim_{\rho\to\infty} \frac{B_{P_X,\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{D_{P_X,\boldsymbol{l}}}{m}. \tag{49}$$

Proof: See Appendix F. □

It follows from Theorems 5–7 that, at high SNR, the BICM-GMI, $M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho)$, and the BEP behave as

$$I^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho) \approx H_{P_X} - \pi D_{P_X,\boldsymbol{l}} Q\left(\sqrt{\rho}d/2\right) \tag{50}$$

$$M^{\mathrm{bicm}}_{P_X,\boldsymbol{l}}(\rho) \approx \frac{\pi d^2}{4} D_{P_X,\boldsymbol{l}} Q\left(\sqrt{\rho}d/2\right) \tag{51}$$

$$B_{P_X,\boldsymbol{l}}(\rho) \approx \frac{D_{P_X,\boldsymbol{l}}}{m} Q\left(\sqrt{\rho}d/2\right). \tag{52}$$

For a given constellation and input distribution, the results in (50)–(52) indicate that, at high SNR, a maximization of the BICM-GMI over binary labelings is equivalent to a minimization of both its derivative and the BEP.

Example 4: Consider the constellation $\mathcal{X} = \{\pm 4, \pm 2\}$ in Example 1 and the labeling $\boldsymbol{l}_{\mathrm{GC}} = [0,1,3,2]$, which gives $A_{\mathcal{X}} = C_{\mathcal{X},\boldsymbol{l}} = 4$. Furthermore, consider the three input distributions

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

which are induced by the bit probabilities listed in the second column of Table II. Table II further lists H_{P_X}, $D_{P_X,\boldsymbol{l}}$, and d

[5]The BEP in (46) is based on hard-decisions made by the *symbol-wise* MAP demapper. Alternatively, one could study a *bit-wise* MAP demapper for which $\hat{Q}_k^{\mathrm{map}}(y) = \mathrm{argmax}_{b\in\{0,1\}} P_{Q_k|Y}(b|y)$. This demapper minimizes the BEP [34] (see also [35]), but its analysis is much more involved.

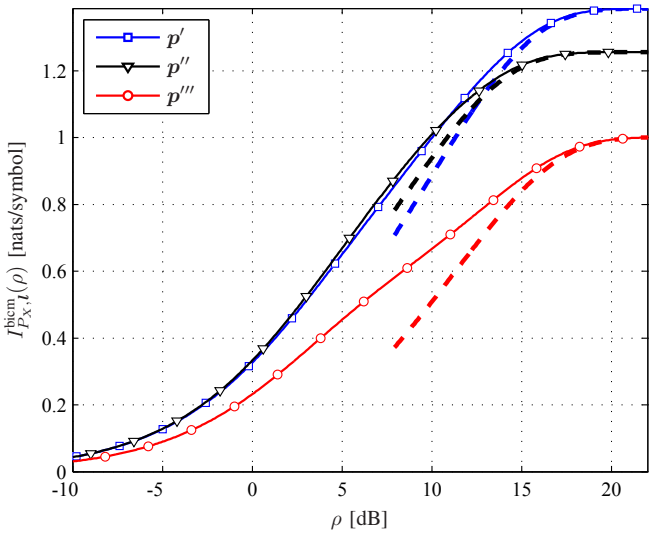

Fig. 5. $I^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho)$ for the three input distributions in Example 4 and the constellation $\mathcal{X} = \{\pm 4, \pm 2\}$ (normalized to $E_{\text{s}} = 1$) (solid lines with markers) and the asymptotic expression in (50) (thick dashed lines).

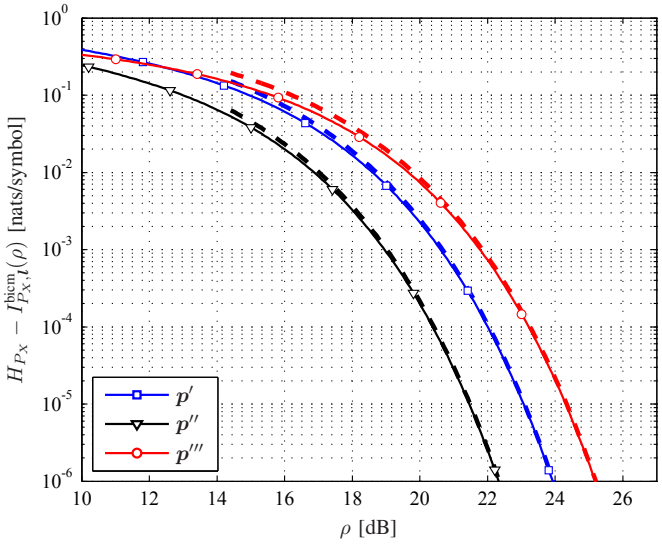

Fig. 6. $H_{P_X} - I^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho)$ for the three input distributions in Example 4 and the constellation $\mathcal{X} = \{\pm 4, \pm 2\}$ (normalized to $E_{\text{s}} = 1$) (solid lines with markers) and the asymptotic expression in (50) (thick dashed lines), i.e., $H_{P_X} - I^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho) \approx \pi D_{P_X,\boldsymbol{l}} Q\left(\sqrt{\rho} d/2\right)$.

when the constellation is normalized to $E_{\text{s}} = 1$. Fig. 5 shows the BICM-GMI curves and Fig. 6 shows the corresponding curves for $H_{P_X} - I^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho)$. The asymptotic expression (50) is also shown. Observe how this asymptotic expression approximates well the BICM-GMI for a large range of SNR.

For a uniform input distribution, Theorems 5–7 particularize to the following result.

Corollary 8: For any $\mathcal{X}$ and $\boldsymbol{l}$ and a uniform input distribution

$$\lim_{\rho\to\infty} \frac{\log M - I^{\text{bicm}}_{P_X^{\text{u}},\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho} d/2\right)} = \pi \frac{C_{\mathcal{X},\boldsymbol{l}}}{M} \tag{53}$$

$$\lim_{\rho\to\infty} \frac{M^{\text{bicm}}_{P_X^{\text{u}},\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho} d/2\right)} = \frac{\pi d^2}{4} \frac{C_{\mathcal{X},\boldsymbol{l}}}{M} \tag{54}$$

$$\lim_{\rho\to\infty} \frac{B_{P_X^{\text{u}},\boldsymbol{l}}(\rho)}{Q\left(\sqrt{\rho} d/2\right)} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{mM} \tag{55}$$

where $C_{\mathcal{X},\boldsymbol{l}}$ is given by (33).

Proof: From Theorems 5–7 and (38). □

The expression in (55) for the BEP is well-known, see, e.g., [36, p. 130]. The asymptotic results for BICM are summarized in Table III.

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

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

D. Classification of Labelings at high SNR

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

$$K^{\text{mi}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \frac{H_{P_X} - I^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho)}{H_{P_X} - I_{P_X}(\rho)} \tag{56}$$

$$K^{\text{mmse}}_{P_X,\boldsymbol{l}}(\rho) \triangleq \frac{M^{\text{bicm}}_{P_X,\boldsymbol{l}}(\rho)}{M_{P_X}(\rho)}. \tag{57}$$

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

$$K^{\text{mi}}_{P_X,\boldsymbol{l}}(\rho) \ge 1. \tag{58}$$

We further define

$$T_{P_X,\boldsymbol{l}} \triangleq \lim_{\rho\to\infty} K^{\text{mi}}_{P_X,\boldsymbol{l}}(\rho) \tag{59}$$

$$= \lim_{\rho\to\infty} K^{\text{mmse}}_{P_X,\boldsymbol{l}}(\rho) \tag{60}$$

where (60) follows from L'Hôpital's rule. Theorems 1 and 5 yield

$$T_{P_X,\boldsymbol{l}} = \frac{D_{P_X,\boldsymbol{l}}}{B_{P_X}}. \tag{61}$$

Furthermore, by (58),

$$T_{P_X,\boldsymbol{l}} \ge 1. \tag{62}$$

We next study $T_{P_X,\boldsymbol{l}}$ for a uniform input distribution P_X^{u}. With a slight abuse of notation, we will refer to $T_{P_X^{\text{u}},\boldsymbol{l}}$ as $T_{\mathcal{X},\boldsymbol{l}}$.

Corollary 9: For any labeling $\boldsymbol{l}$ and constellation $\mathcal{X}$,

$$T_{\mathcal{X},\boldsymbol{l}} = \frac{C_{\mathcal{X},\boldsymbol{l}}}{A_{\mathcal{X}}}. \tag{63}$$

Proof: Follows by using (38) and (15) in (61). □

By Corollary 9, an upper bound on $C_{\mathcal{X},\boldsymbol{l}}$ yields an upper bound on $T_{\mathcal{X},\boldsymbol{l}}$.

Theorem 10: For any one-dimensional constellation $\mathcal{X}$ and any labeling $\boldsymbol{l}$

$$C_{\mathcal{X},\boldsymbol{l}} \leq \min\left(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M\right) \tag{64}$$

and hence

$$T_{\mathcal{X},\boldsymbol{l}} \leq \frac{\min\left(mA_{\mathcal{X}}, (m-1)A_{\mathcal{X}} + M\right)}{A_{\mathcal{X}}}. \tag{65}$$

Proof: By definition, we have $A_{\mathcal{X}_{k,0}} \geq 0$ and $A_{\mathcal{X}_{k,1}} \geq 0$ which by (34) yields

$$C_{\mathcal{X},\boldsymbol{l}} \leq mA_{\mathcal{X}}. \tag{66}$$

This bound holds with equality if the labels of all $A_{\mathcal{X}}/2$ pairs of constellation points at MED differ in exactly m bits. Conversely, (66) can only hold with equality if $A_{\mathcal{X}} \leq M$, since there are only $M/2$ pairs of labels at Hamming distance m. For $A_{\mathcal{X}} > M$, the quantity $C_{\mathcal{X},\boldsymbol{l}}$ is maximized if the labels of $M/2$ constellation pairs differ in m bits and the labels of the remaining $(A_{\mathcal{X}} - M)/2$ pairs differ in $m-1$ bits, which gives

$$C_{\mathcal{X},\boldsymbol{l}} \leq mM + (m-1)(A_{\mathcal{X}} - M) \tag{67}$$

$$= (m-1)A_{\mathcal{X}} + M, \quad A_{\mathcal{X}} > M. \tag{68}$$

Combining (66) and (68) proves (64), which together with (63) proves (65). □

For an MPAM constellation, Theorem 10 specializes to

$$C_{\mathcal{E},\boldsymbol{l}} \leq 2mM - 2m - M + 2 \tag{69}$$

$$T_{\mathcal{E},\boldsymbol{l}} \leq m - \frac{M-2}{2M-2}. \tag{70}$$

Furthermore, if the MPAM constellation is labeled with the NBC, we obtain from (35)

$$T_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}} = \frac{2M - m - 2}{M-1}. \tag{71}$$

Example 5: In Fig. 7, we show the functions $K^{\mathrm{mi}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ and $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ in (56) and (57), respectively, for a 4PAM constellation with a uniform input distribution ($P_X = P^{\mathrm{eu}}_X$, $A_{\mathcal{X}} = 6$) and the three labelings that give a different BICM-GMI: $\boldsymbol{l}_{\mathrm{GC}} = [0,1,3,2]$, $\boldsymbol{l}_{\mathrm{NBC}} = [0,1,2,3]$, and $\boldsymbol{l}_{\mathrm{AGC}} = [0,3,2,1]$.[6] The values of $T_{\mathcal{E},\boldsymbol{l}}$ are also shown. In contrast to the BICM-GMI curves plotted, e.g., in [17, Fig. 3] and [33, Fig. 1], the functions $K^{\mathrm{mi}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ and $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ allow us to study different labelings at high SNR. Observe that the GC gives $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{GC}}} = 1$, and that the AGC achieves the upper bound in (70), i.e., $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{AGC}}} = 5/3$. The function $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ 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 $M^{\mathrm{bicm}}_{P^{\mathrm{u}}_X,\boldsymbol{l}}(\rho)$ as ρ tends to zero, which agrees with [20], [21, Theorem 14].[7]

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

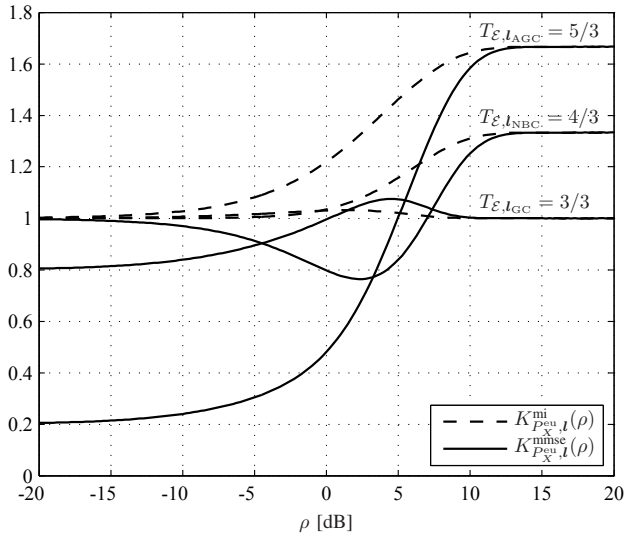

Fig. 7. Functions $K^{\mathrm{mi}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ (dashed lines) and $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ (solid lines) for 4PAM (normalized to $E_{\mathrm{s}} = 1$) and all three nonequivalent labelings. The values of $T_{\mathcal{E},\boldsymbol{l}}$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

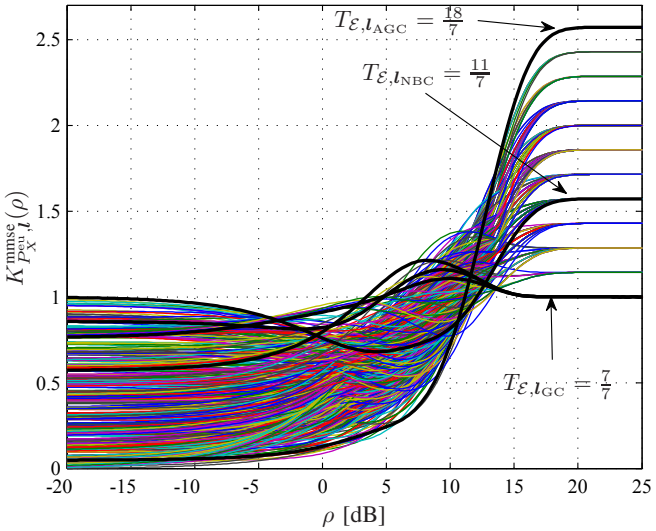

Fig. 8. $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ for the 458 labelings that give a different BICM-GMI for 8PAM (normalized to $E_{\mathrm{s}} = 1$). The values of $T_{\mathcal{E},\boldsymbol{l}}$ for the three nonequivalent GCs, the NBC, and the AGC are also shown (thick lines).

Example 6: In Fig. 8, we show the function $K^{\mathrm{mmse}}_{P^{\mathrm{eu}}_X,\boldsymbol{l}}(\rho)$ for 8PAM ($P_X = P^{\mathrm{eu}}_X$, $A_{\mathcal{X}} = 14$) and all the 458 labelings that give a different BICM-GMI [23]. The value $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}}$ obtained using (71) is also shown. We further highlight the three nonequivalent GCs (in terms of BEP) [31, 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 $\boldsymbol{l} = [0,1,3,7,5,4,6,2]$. All these GCs give $T_{\mathcal{E},\boldsymbol{l}} = 1$. Observe that there are 12 possible values of $T_{\mathcal{E},\boldsymbol{l}}$, which is consistent

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

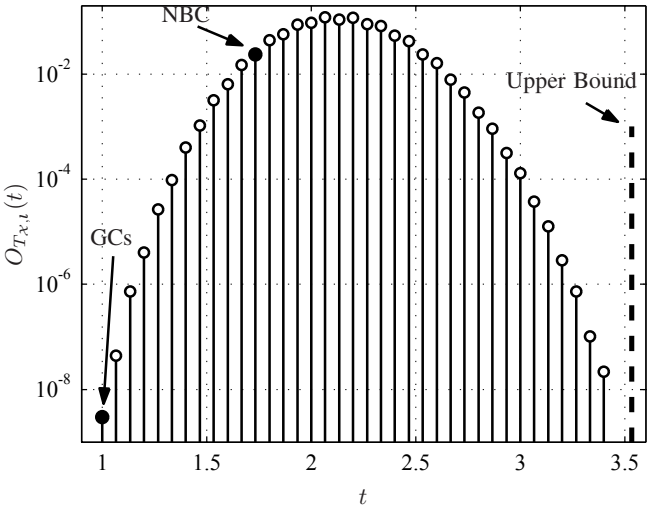

Fig. 9. Approximated $O_{T_{\mathcal{X},\boldsymbol{l}}}(t)$ using 10^9 randomly generated labelings for 16PAM (normalized to $E_{\mathrm{s}}=1$). For GCs, we have $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{GC}}}=1$ and for the NBC, we have $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}}=26/15$. The upper bound (70) is also shown.

with [23, Fig. 3].[8] Using (55), the 12 values of $T_{\mathcal{E},\boldsymbol{l}}$ in Fig. 8 translate into 12 different asymptotic BEP curves, which were recently reported in [35, Fig. 4].

Example 7: Motivated by [21, Fig. 6], we study the relative occurrence of labelings with a given $T_{\mathcal{X},\boldsymbol{l}}$, i.e.,

$$O_{T_{\mathcal{X},\boldsymbol{l}}}(t) \triangleq \frac{\text{number of labelings with } T_{\mathcal{X},\boldsymbol{l}}=t}{\text{total number of labelings}}. \tag{72}$$

In Fig. 9, we present an approximation of $O_{T_{\mathcal{X},\boldsymbol{l}}}(t)$ for 16PAM, obtained by randomly generating 10^9 labelings. We highlight $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{GC}}}$ and $T_{\mathcal{E},\boldsymbol{l}_{\mathrm{NBC}}}$. The upper bound (70) is also shown. Observe that most of the possible labelings are not Gray.

E. Gray Codes, Anti-Gray Codes, and Asymptotic Optimality

In view of the lower bound (62), we say that a labeling $\boldsymbol{l}$ is *asymptotically optimal* (AO) in terms of BICM-GMI for a constellation $\mathcal{X}$ and a uniform input distribution if it satisfies $T_{\mathcal{X},\boldsymbol{l}}=1$. 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 (71), 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 the conjecture of the optimality of GCs at high SNR in terms of BICM-GMI [14, Sec. III-C].

Theorem 11: For any constellation $\mathcal{X}$ and a uniform input distribution, a labeling is AO if and only if it is a GC.

Proof: By definition, for a GC, all pairs of labelings of constellation points at MED are at Hamming distance one. Thus, $A_{\mathcal{X}}=C_{\mathcal{X},\boldsymbol{l}}$, and by (63), $T_{\mathcal{X},\boldsymbol{l}}=1$, demonstrating that every GC is AO. Conversely, for every non-GC, there is at least one pair of constellation points at MED with Hamming distance larger than one. Consequently, every non-GC gives $C_{\mathcal{X},\boldsymbol{l}} > A_{\mathcal{X}}$, and therefore, $T_{\mathcal{X},\boldsymbol{l}} > 1$. □

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

Remark 3: The optimality of GCs directly extends 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 studied 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.

Theorem 11 shows that GCs minimize $T_{\mathcal{X},\boldsymbol{l}}$. In what follows, we show that, for MPAM constellations, it is always possible to construct a labeling that maximizes $T_{\mathcal{E},\boldsymbol{l}}$, i.e., a labeling that achieves the upper bound (70).

Let $\mathcal{C}_{\mathcal{X}}$ denote the set of all possible values that $C_{\mathcal{X},\boldsymbol{l}}$ can take. Noting that $C_{\mathcal{X},\boldsymbol{l}}$ is an even integer bounded by (36) and (64), it follows that, for any constellation $\mathcal{X}$, the cardinality of $\mathcal{C}_{\mathcal{X}}$ satisfies

$$|\mathcal{C}_{\mathcal{X}}| \leq \frac{1}{2}\min\left\{(m-1)A_{\mathcal{X}}+2, (m-2)A_{\mathcal{X}}+M+2\right\}. \tag{73}$$

The expression (73) 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 $\mathcal{X}=\mathcal{E}$, we obtain from (13) and (73)

$$|\mathcal{C}_{\mathcal{E}}| \leq mM - \frac{3M}{2} - m + 3. \tag{74}$$

For example, for 4PAM we have $|\mathcal{C}_{\mathcal{E}}| \leq 3$ and for 8PAM we have $|\mathcal{C}_{\mathcal{E}}| \leq 12$, which is consistent with the 3 and 12 classes at high SNR shown in Fig. 7 and Fig. 8, respectively. For 16PAM, the upper bound (74) indicates that there are at most 39 classes. However, Fig. 9 shows only 37 classes, all giving rise to a $T_{\mathcal{E},\boldsymbol{l}}$ strictly smaller than (70). This raises the question of whether to produce Fig. 9 we were drawing not enough labelings[9] or whether the upper bounds (70) and (74) are loose. As we shall show next, (70) is achieved by an AGC.

The AGC of order $m \geq 2$ is defined by the $M \times m$ binary matrix $\mathbf{W}_m$ (the ith row is the binary label for x_i) where $\mathbf{W}_m$ is constructed according to the following recursive procedure:

Let $\mathbf{W}_1 = [0,1]^{\mathsf{T}}$. Construct $\mathbf{W}_m$ from $\mathbf{W}_{m-1}$ following the next three steps:

Step 1 Reverse the order of 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 $\mathbf{W}_m$.

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

Lemma 12: All the rows in $\mathbf{W}_m$ are unique and every odd row differs in m bits compared to its subsequent row.

[9]Without discarding trivial operations, there are $16! \approx 2.1 \cdot 10^{13}$ labelings, so randomly generating 10^9 labelings covers only a small fraction of all possible labelings.

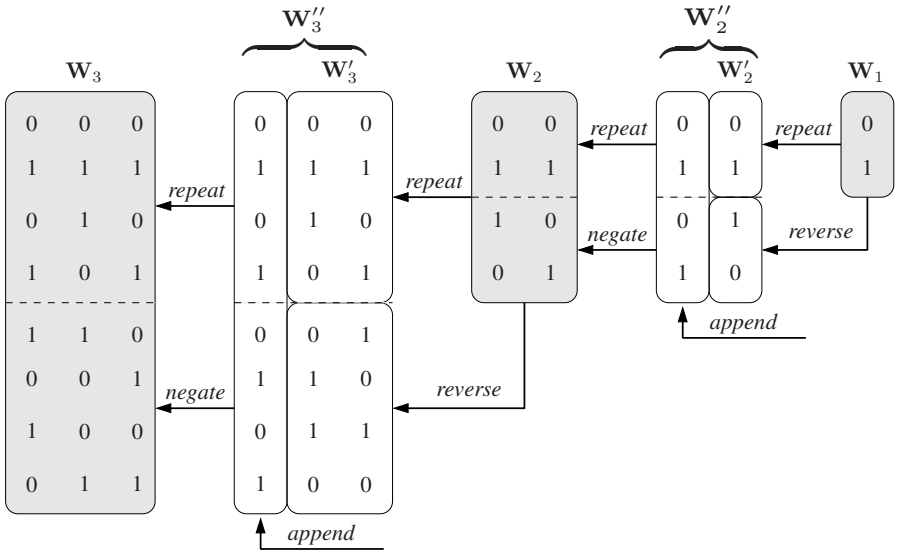

Fig. 10. Proposed recursive construction of an AGC for $m = 2$ and $m = 3$.

Proof: We shall prove Lemma 12 by induction. Let all rows in $\mathbf{W}_{m-1}$ be unique and every odd row differ in $m-1$ bits compared to its subsequent row. $\mathbf{W}_1$ clearly fulfills both criteria. Since the number of rows in $\mathbf{W}_{m-1}$ is even, it follows by Step 1 that 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. This 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. Negating 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. Consequently, all rows in $\mathbf{W}_m$ are unique and every odd row differs in m bits compared to its subsequent row. □

We next show that, for MPAM constellations, the AGC maximizes $T_{\mathcal{E},\boldsymbol{\imath}}$.

Theorem 13: For $\mathcal{X} = \mathcal{E}$, the AGC achieves the upper bounds in (69) and (70), i.e.,

$$C_{\mathcal{E},\boldsymbol{\imath}_{\mathrm{AGC}}} = 2mM - 2m - M + 2 \tag{75}$$

$$T_{\mathcal{E},\boldsymbol{\imath}_{\mathrm{AGC}}} = m - \frac{M-2}{2M-2}. \tag{76}$$

Proof: Let H_m 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. Recall that every MPAM constellation satisfies $H_m = C_{\mathcal{E},\boldsymbol{\imath}}$. 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 $H_m = 2\left(mM - m - M/2 + 1\right)$, which proves (75). Together with (63) and (13), this proves Theorem 13. □

The labeling $\boldsymbol{\imath}_{\mathrm{AGC}} = [0,3,2,1]$ in Example 5 and Fig. 7 (i.e., $\mathbf{W}_2$ in Fig. 10) is the AGC for 4PAM and gives $T_{\mathcal{E},\boldsymbol{\imath}} = 5/3$. For 8PAM, the AGC is $\boldsymbol{\imath}_{\mathrm{AGC}} = [0,7,2,5,6,1,4,3]$ ($\mathbf{W}_3$ in Fig. 10) and gives $T_{\mathcal{E},\boldsymbol{\imath}} = 18/7$ as shown in Fig. 8.

Revisiting Example 7, we note that, by Theorem 13, the labeling that achieves the upper bound (70) is the AGC $\mathbf{W}_4$. It can be further shown that the labeling with the second largest $T_{\mathcal{E},\boldsymbol{\imath}}$ can be constructed by reversing the order of the three first rows of the AGC $\mathbf{W}_4$. This demonstrates that for 16PAM there exist indeed 39 classes of labelings with different high-SNR behaviors and hence the bound in (74) is tight for this case.

V. Conclusions

We studied the discrete-time, scalar (real-valued) AWGN channel when the input takes value in a finite constellation and derived high-SNR asymptotic expressions for the MI, MMSE, SEP, the BICM-GMI, its derivative, and the BEP. Our results show that, as the SNR tends to infinity, all these quantities converge to their asymptotes proportionally to $Q\left(\sqrt{\rho}d/2\right)$, where d is the MED of the constellation.

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. Consequently, 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 study binary labelings for BICM. By characterizing the high-SNR behavior of the BICM-GMI, we proved the long-standing conjecture that Gray codes are optimal at high SNR. 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

Both the numerator and the denominator on the left-hand side (l.h.s.) of (17) tend to zero as ρ tends to infinity. Thus, it

follows from L'Hôpital's rule that

$$\lim_{\rho\to\infty}\frac{H_{P_X}-I_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)}=\lim_{\rho\to\infty}\frac{\frac{\mathrm{d}}{\mathrm{d}\rho}\left(H_{P_X}-I_{P_X}(\rho)\right)}{\frac{\mathrm{d}}{\mathrm{d}\rho}Q\left(\sqrt{\rho}d/2\right)} \tag{77}$$

$$=\frac{4}{d^2}\lim_{\rho\to\infty}\frac{M_{P_X}(\rho)}{G\left(\sqrt{\rho}d/2\right)} \tag{78}$$

$$=\frac{4}{d^2}\lim_{\rho\to\infty}\frac{M_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} \tag{79}$$

$$=\pi B_{P_X} \tag{80}$$

where $G(x)$ is defined as

$$G(x)\triangleq\frac{1}{x}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}. \tag{81}$$

Here the last step follows from Theorem 2 (proved in Appendix B), which also demonstrates that the limit on the right-hand side (r.h.s.) of (77) exists. To pass from (77) to (78) we used (16) and

$$\frac{\mathrm{d}}{\mathrm{d}\rho}Q\left(\sqrt{\rho}d/2\right)=-\frac{d^2}{8}G\left(\sqrt{\rho}d/2\right). \tag{82}$$

To pass from (78) to (79) we used [37, Prop. 19.4.2] to obtain

$$\lim_{x\to\infty}\frac{G(x)}{Q(x)}=1. \tag{83}$$

This proves Theorem 1.

Appendix B
Proof of Theorem 2

For the AWGN channel in (1), the conditional ME is given by [5, eq. (22)]

$$\hat{X}^{\mathrm{me}}(y)=\frac{\sum_{j\in\mathcal{I}_{\mathcal{X}}}p_jx_j\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}{\sum_{j\in\mathcal{I}_{\mathcal{X}}}p_j\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}. \tag{84}$$

By using (84) in (7), we obtain

$$M_{P_X}(\rho)=\sum_{i\in\mathcal{I}_{\mathcal{X}}}p_i\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_i)^2}\cdot\left(\frac{\sum_{j\in\mathcal{I}_{\mathcal{X}}}p_j(x_i-x_j)\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}{\sum_{j\in\mathcal{I}_{\mathcal{X}}}p_j\mathrm{e}^{-\frac{1}{2}(y-\sqrt{\rho}x_j)^2}}\right)^2\mathrm{d}y \tag{85}$$

$$=\sum_{i\in\mathcal{I}_{\mathcal{X}}}p_iV_i(\rho) \tag{86}$$

where

$$V_i(\rho)\triangleq\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-t^2}}{\sqrt{\pi}}\left(\frac{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}}\delta R_{P_X}^{(i)}(\delta)\cdot\mathrm{e}^{-\sqrt{2\rho}t\delta-\frac{\rho\delta^2}{2}}}{\sum_{\delta\in\mathcal{D}_{\mathcal{X}}^{(i)}}R_{P_X}^{(i)}(\delta)\cdot\mathrm{e}^{-\sqrt{2\rho}t\delta-\frac{\rho\delta^2}{2}}}\right)^2\mathrm{d}t \tag{87}$$

with $\mathcal{D}_{\mathcal{X}}^{(i)}\triangleq\{\delta:\delta=x_i-x,x\in\mathcal{X}\}$ and

$$R_{P_X}^{(i)}(\delta)\triangleq\begin{cases}\frac{p_j}{p_i}, & \text{if } \exists x_j\in\mathcal{X}:x_i-x_j=\delta\\ 0, & \text{otherwise}\end{cases} \tag{88}$$

and where to pass from (85) to (86) we used the substitution $y-\sqrt{\rho}x_i=\sqrt{2}t$.

Combining (83) and (86), we obtain

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

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

Using (87) and (81), and the substitution $r=d\sqrt{\rho/8}$, 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

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

and

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

We will next calculate the first limit in (90). Using the substitution $t=u/r-r$ we express $F_i^-(r)$ as

$$F_i^-(r)=\int_{-\infty}^{\infty}f_i^-(r,u)\,\mathrm{d}u \tag{93}$$

where

$$f_i^-(r,u)\triangleq h(r^2-u)\cdot\mathrm{e}^{2u-\frac{u^2}{r^2}}\cdot\left(\frac{\sum_{\delta\in\mathcal{D}_i^*}\delta R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{d}-4r^2U(\delta)}}{1+\sum_{\delta\in\mathcal{D}_i^*}R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{d}-4r^2U(\delta)}}\right)^2 \tag{94}$$

with $\mathcal{D}_i^*\triangleq\mathcal{D}_{\mathcal{X}}^{(i)}\setminus\{0\}$,

$$U(\delta)\triangleq\frac{\delta}{d}\left(\frac{\delta}{d}-1\right) \tag{95}$$

and $h(x)$ being Heaviside's step function (i.e., $h(x)=1$ if $x\geq 0$ and $h(x)=0$ if $x<0$). Using the fact that $U(\delta)\geq 0,\forall\delta\in\mathcal{D}_i^*$ and $U(d)=0$, we obtain

$$\lim_{r\to\infty}f_i^-(r,u)=d^2\mathrm{e}^{2u}\left(\frac{R_{P_X}^{(i)}(d)\mathrm{e}^{-4u}}{1+R_{P_X}^{(i)}(d)\mathrm{e}^{-4u}}\right)^2. \tag{96}$$

As we shall prove in Lemma 14 ahead, $u\mapsto f_i^-(r,u)$ is uniformly bounded by some integrable function $u\mapsto g_i^-(u)$ that is independent of r. It thus follows from Lebesgue's

Dominated Convergence Theorem [38, Theorem 1.34] that

$$\lim_{r\to\infty} F_i^-(r) = \int_{-\infty}^{\infty} \lim_{r\to\infty} f_i^-(r,u)\,\mathrm{d}u \tag{97}$$

$$= \frac{d^2\sqrt{R_{P_X}^{(i)}(d)}}{2}\int_0^{\infty} \frac{\xi^2}{(1+\xi^2)^2}\,\mathrm{d}\xi \tag{98}$$

$$= \frac{d^2\pi\sqrt{R_{P_X}^{(i)}(d)}}{8} \tag{99}$$

where (98) follows from (96) and the substitution $\sqrt{R_{P_X}^{(i)}(d)}\mathrm{e}^{-2u} = \xi$, and (99) follows from [39, eq. (3.241.5)].

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

Lemma 14: For any $r > 0$

$$0 \le f_i^-(r,u) \le g_i^-(u), \quad u \in \mathbb{R} \tag{100}$$

where

$$g_i^-(u) \triangleq \frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{101}$$

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

$$\int_{-\infty}^{\infty} g_i^-(u)\,\mathrm{d}u = \frac{\hat{d}^2(M-1)^2}{p_i^2} < \infty. \tag{102}$$

Proof: We first note that $f_i^-(r,u) \ge 0$, $r > 0$, $u \in \mathbb{R}$. It thus remains to show the second inequality in (100). To this end, we use $h(r^2-u) \le 1$, $\mathrm{e}^{-\frac{u^2}{r^2}} \le 1$, and $\delta \le \hat{d}$ to upper-bound (94) as

$$f_i^-(r,u) \le \hat{d}^2\mathrm{e}^{2u}\left(\frac{\sum_{\delta\in\mathcal{D}_i^*} R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}{1+\sum_{\delta\in\mathcal{D}_i^*} R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}\right)^2 \tag{103}$$

$$= \hat{d}^2\mathrm{e}^{2u}\left(1+\frac{1}{\sum_{\delta\in\mathcal{D}_i^*} R_{P_X}^{(i)}(\delta)\mathrm{e}^{-4u\frac{\delta}{\hat{d}}-4r^2U(\delta)}}\right)^{-2}. \tag{104}$$

Since $R_{P_X}^{(i)}(\delta) < 1/p_i$ and $\mathrm{e}^{-4r^2U(\delta)} \le 1$, we can further upper-bound (104) as

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

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

where to pass from (105) to (106) we used $p_i < 1$.

For $u \ge 0$, we have

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

$$< \frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-6u} \tag{108}$$

$$< \frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{109}$$

where to pass from (106) to (107) we upper-bounded the $(M-1)$ exponentials in the summation by e^{-4u}.

For $u \le 0$, we have

$$f_i^-(r,u) < \frac{\hat{d}^2\mathrm{e}^{2u}}{p_i^2} \tag{110}$$

$$\le \frac{\hat{d}^2(M-1)^2}{p_i^2}\mathrm{e}^{-2|u|} \tag{111}$$

where (110) follows from discarding the sum of exponentials in (106). Combining (109) and (111) gives (101). This proves Lemma 14. □

Returning to the proof of Theorem 2, the second limit on the r.h.s. of (90) can be computed along the same lines by using the substitution $t = u/r + r$ in (92):

$$\lim_{r\to\infty} F_i^+(r) = \frac{d^2\pi\sqrt{R_{P_X}^{(i)}(-d)}}{8}. \tag{112}$$

Using (99) and (112) in (90), and combining the result with (89) yields

$$\lim_{\rho\to\infty}\frac{M_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i\frac{\pi d^2}{4}\left(\sqrt{R_{P_X}^{(i)}(d)}+\sqrt{R_{P_X}^{(i)}(-d)}\right) \tag{113}$$

which in view of (88) and (14) is equal to $\pi d^2 B_{P_X}/4$. This proves Theorem 2.

Appendix C
Proof of Theorem 3

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

$$\hat{X}^{\mathrm{map}}(y) = \operatorname*{argmax}_{x\in\mathcal{X}}\left\{f_{Y|X}(y|x)P_X(x)\right\} \tag{114}$$

$$= x_j, \quad \text{if } y \in \mathcal{Y}_j(\rho) \tag{115}$$

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

$$\mathcal{Y}_j(\rho) \triangleq \{y \in \mathbb{R} : \beta_{j-1}(\rho) \le y < \beta_j(\rho)\} \tag{116}$$

where $\beta_\ell(\rho)$ with $\ell = 0,\ldots,M$ are the $M+1$ thresholds defining the M regions, i.e.,

$$\beta_\ell(\rho) = \begin{cases} -\infty, & \ell = 0 \\ \frac{\log(p_\ell/p_{\ell+1})}{\sqrt{\rho}(x_{\ell+1}-x_\ell)} + \frac{\sqrt{\rho}(x_{\ell+1}+x_\ell)}{2}, & \ell = 1,\ldots,M-1 \\ +\infty, & \ell = M \end{cases} \tag{117}$$

which are obtained by solving

$$p_\ell f_{Y|X}(\beta_\ell(\rho)|x_\ell) = p_{\ell+1}f_{Y|X}(\beta_\ell(\rho)|x_{\ell+1}). \tag{118}$$

The following lemma will be used in this proof as well as in the proof of Theorem 7 (Appendix F).

Lemma 15: For any P_X and $i \in \mathcal{I}_{\mathcal{X}}$

$$\lim_{\rho\to\infty} \frac{Q\left(|\beta_\ell(\rho) - \sqrt{\rho}x_i|\right)}{Q\left(\sqrt{\rho}d/2\right)} = \begin{cases} \sqrt{R^{(i)}_{P_X}(d)}, & \text{if } \ell = i-1 \\ \sqrt{R^{(i)}_{P_X}(-d)}, & \text{if } \ell = i \\ 0, & \text{if } \ell \notin \{i-1, i\} \end{cases} \tag{119}$$

where $\beta_\ell(\rho)$ is given by (117) and $R^{(i)}_{P_X}(\delta)$ by (88).

Proof: We use (117) to obtain

$$\beta_\ell(\rho) - \sqrt{\rho}x_i = \frac{\log(p_\ell/p_{\ell+1})}{\sqrt{\rho}(x_{\ell+1} - x_\ell)} + \frac{\sqrt{\rho}\epsilon_{i,\ell}}{2} \tag{120}$$

where for any i, ℓ

$$\epsilon_{i,\ell} \triangleq x_{\ell+1} + x_\ell - 2x_i. \tag{121}$$

We form the ratio

$$\frac{G\left(|\beta_\ell(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)} = \frac{\rho d(x_{\ell+1} - x_\ell)}{|2\log(p_\ell/p_{\ell+1}) + \rho\epsilon_{i,\ell}(x_{\ell+1} - x_\ell)|} \cdot \exp\left(-\frac{\left(\log\frac{p_\ell}{p_{\ell+1}}\right)^2}{2\rho(x_{\ell+1} - x_\ell)^2} - \frac{\epsilon_{i,\ell}\log\frac{p_\ell}{p_{\ell+1}}}{2(x_{\ell+1} - x_\ell)} - \frac{\rho(\epsilon^2_{i,\ell} - d^2)}{8}\right) \tag{122}$$

and study the limit

$$\lim_{\rho\to\infty} \frac{G\left(|\beta_\ell(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)}. \tag{123}$$

To this end, we distinguish between three cases:

(i) If $i = \ell$ and $x_{\ell+1} - x_\ell = d$, then $\epsilon_{i,\ell} = x_{\ell+1} - x_\ell = d$ and the limit in (123) is $\mathrm{e}^{-\log(p_\ell/p_{\ell+1})/2} = \sqrt{p_{\ell+1}/p_\ell}$.

(ii) If $i = \ell + 1$ and $x_{\ell+1} - x_\ell = d$, then $\epsilon_{i,\ell} = x_\ell - x_{\ell+1} = -d$ and the limit in (123) is $\sqrt{p_\ell/p_{\ell+1}}$.

(iii) In all other cases, $|\epsilon_{i,\ell}| > d$ and the limit in (123) is zero.

A slight change in notation then yields

$$\lim_{\rho\to\infty} \frac{G\left(|\beta_\ell(\rho) - \sqrt{\rho}x_i|\right)}{G\left(\sqrt{\rho}d/2\right)} = \begin{cases} \sqrt{\frac{p_{i+1}}{p_i}}, & \text{if } \ell = i \text{ and } x_{\ell+1} - x_\ell = d \\ \sqrt{\frac{p_{i-1}}{p_i}}, & \text{if } \ell = i-1 \text{ and } x_{\ell+1} - x_\ell = d \\ 0, & \text{otherwise} \end{cases}. \tag{124}$$

Combining (124) with (88) and (83) proves Lemma 15. □

Returning to the proof of Theorem 3, using (115) and (116), the SEP in (8) can be written as

$$S_{P_X}(\rho) = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \Pr\{Y \notin \mathcal{Y}_i(\rho) | X = x_i\} \tag{125}$$

$$= \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \Big(Q\left(\beta_i(\rho) - \sqrt{\rho}x_i\right) + Q\left(\sqrt{\rho}x_i - \beta_{i-1}(\rho)\right)\Big) \tag{126}$$

which gives

$$\lim_{\rho\to\infty} \frac{S_{P_X}(\rho)}{Q(\sqrt{\rho}d/2)} = \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \Bigg(\lim_{\rho\to\infty} \frac{Q\left(\beta_i(\rho) - \sqrt{\rho}x_i\right)}{Q(\sqrt{\rho}d/2)} + \lim_{\rho\to\infty} \frac{Q\left(\sqrt{\rho}x_i - \beta_{i-1}(\rho)\right)}{Q(\sqrt{\rho}d/2)}\Bigg) \tag{127}$$

$$= \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \left(\sqrt{R^{(i)}_{P_X}(-d)} + \sqrt{R^{(i)}_{P_X}(d)}\right) \tag{128}$$

where to pass from (127) to (128) we used Lemma 15 twice, observing that the arguments of both Q-functions are positive for large enough ρ. The proof of Theorem 3 is completed by combining (128) with (14) and (88).

Appendix D
Proof of Theorem 5

Using the expression for the BICM-GMI (40), we have

$$H_{P_X} - I^{\text{bicm}}_{P_X,\boldsymbol{\ell}}(\rho) = \sum_{k=1}^{m}(H_{P_X} - I_{P_X}(\rho)) - \sum_{k=1}^{m}\sum_{b\in\{0,1\}} P_{Q_k}(b)(H_{P_{X_{k,b}}} - I_{P_{X_{k,b}}}(\rho)) - (m-1)H_{P_X} + \sum_{k=1}^{m}\sum_{b\in\{0,1\}} P_{Q_k}(b)H_{P_{X_{k,b}}}. \tag{129}$$

The last two terms on the r.h.s. of (129) cancel because

$$\sum_{k=1}^{m}\sum_{b\in\{0,1\}} P_{Q_k}(b)H_{P_{X_{k,b}}} = -\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}} P_{Q_k}(b)P_{X_{k,b}}(x_i)\log P_{X_{k,b}}(x_i) \tag{130}$$

$$= -\sum_{k=1}^{m}\sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \log\frac{p_i}{P_{Q_k}(q_{i,k})} \tag{131}$$

$$= mH_{P_X} + \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \sum_{k=1}^{m}\log P_{Q_k}(q_{i,k}) \tag{132}$$

$$= mH_{P_X} + \sum_{i\in\mathcal{I}_{\mathcal{X}}} p_i \log\prod_{k=1}^{m} P_{Q_k}(q_{i,k}) \tag{133}$$

$$= mH_{P_X} - H_{P_X} \tag{134}$$

where to pass from (130) to (131) we used (31), and to pass from (133) to (134) we used (30).

We divide both sides of (129) by $Q(\sqrt{\rho}d/2)$ 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

$$\lim_{\rho\to\infty}\frac{H_{P_X}-I^{\text{bicm}}_{P_X,\boldsymbol{\iota}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \pi\sum_{k=1}^{m}\left(B_{P_X}-\sum_{b\in\{0,1\}}P_{Q_k}(b)B_{P_{X_{k,b}}}\right). \quad (135)$$

Theorem 5 follows by showing that the r.h.s. of (135) is equal to $\pi D_{P_X,\boldsymbol{\iota}}$. We shall do this in the following lemma, which will also be used in the proof of Theorem 6.

Lemma 16: We have

$$\sum_{k=1}^{m}\left(B_{P_X}-\sum_{b\in\{0,1\}}P_{Q_k}(b)B_{P_{X_{k,b}}}\right) = D_{P_X,\boldsymbol{\iota}} \quad (136)$$

where B_{P_X} is given by (14), $D_{P_X,\boldsymbol{\iota}}$ by (37),

$$B_{P_{X_{k,b}}} \triangleq \sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}\sum_{\substack{j\in\mathcal{I}_{\mathcal{X}_{k,b}}\\ |x_i-x_j|=d}}\sqrt{P_{X_{k,b}}(x_i)P_{X_{k,b}}(x_j)} \quad (137)$$

and $P_{X_{k,b}}(x)$ is given by (31).

Proof: Express the inner sum on the l.h.s. of (136) using (31) and (137) as

$$\sum_{b\in\{0,1\}}P_{Q_k}(b)B_{P_{X_{k,b}}} = \sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}\sum_{\substack{j\in\mathcal{I}_{\mathcal{X}_{k,b}}\\ |x_i-x_j|=d}}\sqrt{p_jp_i}. \quad (138)$$

Expanding the sum in (14) using $\mathcal{X}=\mathcal{X}_{k,0}\cup\mathcal{X}_{k,1}$, we obtain

$$B_{P_X} = \sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}\sum_{\substack{j\in\mathcal{I}_{\mathcal{X}_{k,b}}\\ |x_i-x_j|=d}}\sqrt{p_jp_i} + 2\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,0}}}\sum_{\substack{j\in\mathcal{I}_{\mathcal{X}_{k,1}}\\ |x_i-x_j|=d}}\sqrt{p_jp_i}. \quad (139)$$

Lemma 16 follows by applying (138) and (139) to the l.h.s. of (136) and by using the definition of $D_{P_X,\boldsymbol{\iota}}$ in (37). □

Appendix E
Proof of Theorem 6

We divide the l.h.s. of (41) and (42) by $Q(\sqrt{\rho}d/2)$ and take the limit as $\rho\to\infty$ to obtain

$$\lim_{\rho\to\infty}\frac{M^{\text{bicm}}_{P_X,\boldsymbol{\iota}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \lim_{\rho\to\infty}\sum_{k=1}^{m}\left(\frac{M_{P_X}(\rho)}{Q\left(\sqrt{\rho}d/2\right)}-\sum_{b\in\{0,1\}}P_{Q_k}(b)\frac{M_{P_{X_{k,b}}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)}\right). \quad (140)$$

Changing the order of summation and limit, and applying Theorem 2 yields

$$\lim_{\rho\to\infty}\frac{M^{\text{bicm}}_{P_X,\boldsymbol{\iota}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{\pi d^2}{4}\sum_{k=1}^{m}\left(B_{P_X}-\sum_{b\in\{0,1\}}P_{Q_k}(b)B_{P_{X_{k,b}}}\right) \quad (141)$$

where $B_{P_{X_{k,b}}}$ is given by (137). Theorem 6 follows by noting that, by Lemma 16, the r.h.s. of (141) is equal to $\frac{\pi d^2}{4}D_{P_X,\boldsymbol{\iota}}$.

Appendix F
Proof of Theorem 7

Using the law of total probability, the BEP in (46) can be written as

$$B_{P_X,\boldsymbol{\iota}}(\rho) = \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\Pr\{\hat{Q}^{\text{map}}_k(Y)\neq q_{k,i}|X=x_i\} \quad (142)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\Pr\left\{Y\in\bigcup_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}}\mathcal{Y}_j(\rho)\middle|X=x_i\right\} \quad (143)$$

with $\mathcal{Y}_j(\rho)$ given by (116) and were we use $\overline{b}$ to denote the negation of a bit b. Using the fact that $\mathcal{Y}_j(\rho)$ are disjoint, we rewrite (143) as

$$B_{P_X,\boldsymbol{\iota}}(\rho) = \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}}\Pr\{Y\in\mathcal{Y}_j(\rho)|X=x_i\} \quad (144)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}}\Gamma_{i,j}(\rho) \quad (145)$$

$$= \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\left(\sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}:j<i}\Gamma_{i,j}(\rho) + \sum_{j\in\mathcal{I}_{\mathcal{X}_{k,\overline{b}}}:j>i}\Gamma_{i,j}(\rho)\right) \quad (146)$$

where

$$\Gamma_{i,j}(\rho) \triangleq Q\left(\beta_{j-1}(\rho)-\sqrt{\rho}x_i\right)-Q\left(\beta_j(\rho)-\sqrt{\rho}x_i\right) \quad (147)$$

$$= Q\left(\sqrt{\rho}x_i-\beta_j(\rho)\right)-Q\left(\sqrt{\rho}x_i-\beta_{j-1}(\rho)\right) \quad (148)$$

and where we have used that $Q(-x)=1-Q(x)$.

By using (147) and (148) in (146), dividing both sides of (146) by $Q\left(\sqrt{\rho}d/2\right)$, and taking the limit as $\rho\to\infty$, we obtain

$$\lim_{\rho\to\infty}\frac{B_{P_X,\boldsymbol{\iota}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{1}{m}\sum_{k=1}^{m}\sum_{b\in\{0,1\}}\sum_{i\in\mathcal{I}_{\mathcal{X}_{k,b}}}p_i\sum_{\ell=1}^{4}s_\ell \quad (149)$$

where

$$s_1 = \sum_{j \in \mathcal{I}_{\mathcal{X}_{k,\overline{b}}} : j<i} \lim_{\rho\to\infty} \frac{Q\left(\sqrt{\rho}x_i - \beta_j(\rho)\right)}{Q\left(\sqrt{\rho}d/2\right)} \tag{150}$$

$$s_2 = -\sum_{j \in \mathcal{I}_{\mathcal{X}_{k,\overline{b}}} : j<i} \lim_{\rho\to\infty} \frac{Q\left(\sqrt{\rho}x_i - \beta_{j-1}(\rho)\right)}{Q\left(\sqrt{\rho}d/2\right)} \tag{151}$$

$$s_3 = \sum_{j \in \mathcal{I}_{\mathcal{X}_{k,\overline{b}}} : j>i} \lim_{\rho\to\infty} \frac{Q\left(\beta_{j-1}(\rho) - \sqrt{\rho}x_i\right)}{Q\left(\sqrt{\rho}d/2\right)} \tag{152}$$

$$s_4 = -\sum_{j \in \mathcal{I}_{\mathcal{X}_{k,\overline{b}}} : j>i} \lim_{\rho\to\infty} \frac{Q\left(\beta_j(\rho) - \sqrt{\rho}x_i\right)}{Q\left(\sqrt{\rho}d/2\right)}. \tag{153}$$

Note that, for sufficiently large ρ, the arguments of the Q-functions in (150)–(153) are positive.

By Lemma 15, $s_2 = s_4 = 0$. Furthermore, applying Lemma 15 to (150) and (152), we conclude that the only nonzero contribution to s_1 and s_3 can come from the terms $j = i-1$ and $j = i+1$, respectively. We therefore express s_1 as

$$s_1 = \begin{cases} \sqrt{\frac{p_{i-1}}{p_i}}, & \text{if } \exists x_{i-1} \in \mathcal{X}_{k,\overline{b}} : x_i - x_{i-1} = d \\ 0, & \text{otherwise} \end{cases} \tag{154}$$

and s_3 as

$$s_3 = \begin{cases} \sqrt{\frac{p_{i+1}}{p_i}}, & \text{if } \exists x_{i+1} \in \mathcal{X}_{k,\overline{b}} : x_{i+1} - x_i = d \\ 0, & \text{otherwise} \end{cases}. \tag{155}$$

Using (154) and (155) in (149), we obtain

$$\lim_{\rho\to\infty} \frac{B_{P_X,\mathcal{U}}(\rho)}{Q\left(\sqrt{\rho}d/2\right)} = \frac{1}{m}\sum_{k=1}^{m} \sum_{b\in\{0,1\}} \sum_{i \in \mathcal{I}_{\mathcal{X}_{k,b}}} p_i \sum_{\substack{j \in \mathcal{I}_{\mathcal{X}_{k,\overline{b}}} \\ |x_i - x_j| = d}} \sqrt{\frac{p_j}{p_i}}. \tag{156}$$

The proof is completed by moving p_i to the inner sum in (156) and by comparing the resulting expression with (37).

ACKNOWLEDGMENT

The authors wish to thank the Associate Editor Robert F. H. Fischer and the anonymous reviewers for their valuable comments.

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. Boutros, and M. Moeneclaey, "Precoding for outage probability minimization on block fading channels," *IEEE Trans. Inf. Theory (to appear)*, 2013, 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*, vol. 59, no. 4, pp. 2396–2410, Apr. 2013.

[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] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of mutual information for discrete constellations," in *IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, July 2013.

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

[29] T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.

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

[31] 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.

[32] 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.

[33] A. Guillén i Fàbregas and A. Martinez, "Derivative of BICM mutual information," *IET Electronics Letters*, vol. 43, no. 22, pp. 1219–1220, Oct. 2007.
[34] M. Ivanov, F. Brännström, A. Alvarado, and E. Agrell, "On the exact BER of bit-wise demodulators for one-dimensional constellations," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1450–1459, Apr. 2013.
[35] ——, "General BER expression for one-dimensional constellations," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Anaheim, CA, Dec. 2012.
[36] U. Madhow, *Fundamentals of Digital Communication*. Cambridge University Press, 2008.
[37] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge University Press, 2009.
[38] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, 1987.
[39] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 6th ed. Academic Press, 2000.

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

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

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

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

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

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

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, first as an Associate Professor and since 2009 as a Professor in Communication Systems. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the signals and systems research area. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications. More specifically, his current research includes polarization-multiplexed coding and modulation, nonlinear channel capacity, optical intensity modulation, cognitive radio, and coded modulation.

Prof. Agrell served as Publications Editor for IEEE TRANSACTIONS ON INFORMATION THEORY from 1999 to 2002 and is an Associate Editor for IEEE TRANSACTIONS ON COMMUNICATIONS since 2012. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, and the 2013 Chalmers Supervisor of the Year Award.

Tobias Koch (S'02, M'09) is a Visiting Professor with the Signal Theory and Communications Department of Universidad Carlos III de Madrid (UC3M), Spain. He received the M.Sc. degree in electrical engineering (with distinction) in 2004 and the Ph.D. degree in electrical engineering in 2009, both from ETH Zurich, Switzerland. From June 2010 until May 2012 he was a Marie Curie Intra-European Research Fellow with the University of Cambridge, UK. He was also a research intern at Bell Labs, Murray Hill, NJ in 2004 and at Universitat Pompeu Fabra (UPF), Spain, in 2007. He joined UC3M in 2012. His research interests include digital communication theory and information theory.

Dr. Koch is serving as Vice Chair of the Spain Chapter of the IEEE Information Theory Society in 2013–2014.