High-SNR Asymptotics of Mutual Information for Discrete Constellations

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

IEEE International Symposium on Information Theory (ISIT) 2013, Istanbul, Turkey

Citation for the published paper:
Alvarado, A. ; Brännström, F. ; Agrell, E. (2013) "High-SNR Asymptotics of Mutual Information for Discrete Constellations". IEEE International Symposium on Information Theory (ISIT) 2013, Istanbul, Turkey

Downloaded from: http://publications.lib.chalmers.se/publication/176337

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

\[ Y = \sqrt{\rho}X + Z \]  

(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

\[ C(\rho) = \frac{1}{2} \log(1 + \gamma) \]  

(2)

where \( \gamma \triangleq \rho E_X [X^2] \) 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]–[3] 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 [3] 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 \( \rho \)) 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 \( \frac{1}{\sqrt{\gamma d/2}} \), 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 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 [15] 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 for low and medium SNRs, there exist other labelings that conjectured in [14, Sec. III-C]; however, it was shown in [17] the binary labeling of the constellation. The optimality of a (suboptimal) bit-wise detection rule, which was cast as 802.11a/g/n, and the DVB standards (DVB-T2/S2/C2).

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 $Q(\sqrt{d}/2)$.

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 $x = [x_1, x_2, \ldots, x_M]$, 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 $\mathbb{B} \triangleq \{0, 1\}$ and the bipolar set as $\mathbb{B} \triangleq \{-1, +1\}$. The negation of a bit $b$ is denoted by $\overline{b}$. All the logarithms are natural logarithms and all the MI s 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 $E_X[\cdot]$.

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 $|\mathcal{X}| = M = 2^n$. The set of indices that enumerates all the constellation symbols in $\mathcal{X}$ is defined as $\mathcal{I}_X \triangleq \{1, \ldots, M\}$.

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 $p = [p_1, \ldots, p_M]$ associated with the symbols. We assume that neither the constellation nor the input distribution depends on $\rho$.

The transmitted average symbol energy is finite and given by

$$E_s \triangleq E_X[X^2] = \sum_{i \in \mathcal{I}_X} p_i x_i^2.$$  (3)

It follows that the SNR $\gamma$ in (2) is $\gamma = \rho E_s$.

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

The Gaussian Q-function is defined as

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

the entropy of the random variable $X$ as

$$H_X(\rho) \triangleq -E_{X,Y} \left[ \log \left( p_{X}(X) \right) \right]$$ (5)

and the MI between $X$ and $Y$ as

$$I_{XY}(\rho) \triangleq E_{X,Y} \left[ \log \left( f_{Y|X}(Y|X)/f_Y(Y) \right) \right]$$ (6)

and the MMSE as

$$M_{XY}(\rho) \triangleq E_{X,Y} \left[ (X - \hat{X}_{\text{MAP}}(Y))^2 \right]$$ (7)

where $\hat{X}_{\text{MAP}}(y) \triangleq \arg \max_{x \in \mathcal{X}} P_{X|Y}(x|y)$ is the conditional (posterior) mean estimator.

We also define the SEP as

$$S_{X}(\rho) \triangleq \text{Pr}\{\hat{X}_{\text{MAP}}(Y) \neq X\}$$ (8)

where $X$ is the transmitted symbol and

$\hat{X}_{\text{MAP}}(y) \triangleq \arg \max_{x \in \mathcal{X}} P_{X|Y}(x|y)$ (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

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

We define the counting function

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

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

$$A_{\mathcal{X}} \triangleq \sum_{i \in \mathcal{I}_{X}} \sum_{w \in \mathcal{W}} A_X^{(i)}(wd).$$ (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
\[
2 \leq A_X \leq 2(M - 1).
\] (13)

The upper bound is achieved by an MPAM constellation, for which
\[
A_E = 2(M - 1).
\] (14)

Analogous to \(A^{(i)}_X(\delta)\), we define \(B^{(i)}_{P_X}(\delta)\) as
\[
B^{(i)}_{P_X}(\delta) = \begin{cases} \sqrt[4]{p_i p_i}, & \text{if } \exists x_j \in X : x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases}
\] (15)

Clearly \(B^{(i)}_{P_X}(0) = p_i, \forall i \in \mathcal{I}_X\).

Finally, for a given \(P_X\), we define the constant
\[
B_{P_X} = \sum_{i \in X} \sum_{w \in W} B^{(i)}_{P_X}(wd).
\] (16)

For a uniform input distribution, \(P_X = P_X^0\) and \(MB_{P_X}^{(i)}(\delta) = A_X^{(i)}(\delta)\), so
\[
B_{P_X} = \frac{A_X}{M}.
\] (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 \(\mathbb{R}^2\) is \(d = 2\), \(E_\mathcal{X}\) in \(\mathbb{C}\) is \(E_\mathcal{X} = 10\), \(A_X = 4\) (two pairs of constellation points at MED), and \(B_{P_X}\) in \(\mathbb{C}\) is \(B_{P_X} = 2\sqrt{p_i p_i} + 2\sqrt{p_i p_i} \approx 0.98\). This example will be continued in Example 3.

### III. High-SNR Asymptotics

There exists a fundamental relationship between the MI and the MMSE for AWGN channels \(\mathbb{C}\) (see also \(\mathbb{R}\) Ch. 2):
\[
\frac{d}{dp} I_{P_X}(\rho) = \frac{1}{2} M_{P_X}^{(i)}(\rho).
\] (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 \([5, 17, 19–12]\). 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 \(\rho\) tends to infinity. In the following we study how fast the MI converges towards its maximum \(H_{P_X}\) by analyzing the difference \(H_{P_X} - I_{P_X}(\rho)\). Theorem 1 is the main result of this paper and characterizes the high-SNR behavior of \(H_{P_X} - I_{P_X}(\rho)\).

**Theorem 1:** For any \(P_X\)
\[
\lim_{\rho \to \infty} \frac{H_{P_X} - I_{P_X}(\rho)}{Q\left(\sqrt{\rho d}/2\right)} = \pi B_{P_X}
\] (19)
where \(B_{P_X}\) is given by \(16\).

**Proof:** The proof is given in Appendix A.

**Example 1:** Theorem 1 we have the following asymptotic expression for the MMSE.

**Theorem 2:** For any \(P_X\)
\[
\lim_{\rho \to \infty} \frac{M_{P_X}(\rho)}{Q\left(\sqrt{\rho d}/2\right)} = \frac{\pi d^2}{4} B_{P_X}
\] (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 \(P_X\)
\[
\lim_{\rho \to \infty} \frac{S_{P_X}(\rho)}{Q\left(\sqrt{\rho d}/2\right)} = B_{P_X}
\] (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
\[
I_{P_X}(\rho) \approx H_{P_X} - \pi B_{P_X} Q\left(\sqrt{\rho d}/2\right),
\] (22)
\[
M_{P_X}(\rho) \approx \frac{\pi d^2}{4} B_{P_X} Q\left(\sqrt{\rho d}/2\right),
\] (23)
\[
S_{P_X}(\rho) \approx B_{P_X} Q\left(\sqrt{\rho d}/2\right).
\] (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, 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 \([22, eq. (1)]\) 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 \((P_X = P_X^0)\), Theorems 1–3 particularize to the following result.

1 The quantity \(H_{P_X} - I_{P_X}(\rho)\) is the conditional entropy of \(X\) given \(Y\).
Corollary 1: For any $\mathcal{X}$ with a uniform input distribution
\[
\lim_{\rho \to \infty} \frac{\log M - I_{\rho_d}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{\pi A_X}{M},
\]
\[
\lim_{\rho \to \infty} \frac{M_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{\pi d^2 A_X}{4M},
\]
\[
\lim_{\rho \to \infty} \frac{S_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{A_X}{M},
\]
where $A_X$ is given in (12).

Proof: From Theorems 1 and 2.

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 $M$-PAM and a uniform input distribution ($P_X = P_X^{\text{uni}}$), Corollary 1 particularizes to (see 14)
\[
I_{P_{\rho_d}}(\rho) \approx \log M - \frac{2\pi(M - 1)}{M} Q(\sqrt{\rho d}/2),
\]
\[
M_{P_X^{\text{uni}}}(\rho) \approx \frac{6\pi E_s}{M(M + 1)} Q(\sqrt{\rho d}/2),
\]
\[
S_{P_X^{\text{uni}}}(\rho) \approx \frac{2(M - 1)}{M} Q(\sqrt{\rho d}/2).
\]

In Table I the results obtained in Theorems 1 and 2 are summarized.

Example 2: In Fig. 1, we show the conditional entropy $\log M - I_{P_{\rho_d}}(\rho)$ for 4PAM and 16PAM with uniform input distributions 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 $I_{P_{\rho_d}}(\rho)$ 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 $A_X$. For one-dimensional constellations with the same $E_s$, the MED is maximized by an $M$-PAM constellation ($\mathcal{X} = \mathcal{E}$).

We conclude this section by noting that if Theorems 1 and 2 are combined, we obtain
\[
\lim_{\rho \to \infty} \frac{M_{P_X}(\rho)}{H_{P_X} - I_{P_{\rho_d}}(\rho)} = \frac{d^2}{4}.
\]
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 2 we obtain
\[
\lim_{\rho \to \infty} \frac{H_{P_X} - I_{P_{\rho_d}}(\rho)}{S_{P_X}(\rho)} = \pi.
\]
Thus, for any $P_X$, the limiting ratio between the conditional entropy and the SEP equals $\pi$.

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 (ENC), a bit-level interleaver ($\Pi$), and a memoryless mapper ($\Phi$). The BICM decoder is based on a demapper ($\Phi^{-1}$) that computes logarithmic likelihood ratios, a de-interleaver ($\Pi^{-1}$), and a channel decoder (DEC).

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

In BICM, the input symbols are possibly nonuniformly distributed. Therefore, the vector of bit positions of the mapper (see Fig. 3) are assumed to be independent but not uniformly distributed. A key element for the performance of BICM is $\Phi: B^m \to X$, 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 SEP. 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 $l = [l_1, l_2, \ldots, l_M]$ where $l_i \in \{0, 1, \ldots, M - 1\}$ is the integer representation of the $i$th length-$m$ binary label $q_i = [q_{i,1}, \ldots, q_{i,m}] \in B^m$ associated with the symbol $x_i$, with $q_{i,1}$ being the most significant bit. The labeling defines $2m$ subconstellations $X_{k,b} \subset X$ for $k = 1, \ldots, m$ and $b \in B$, given by $X_{k,b} = \{x_i \in X : q_{i,k} = b\}$ with $|X_{k,b}| = M/2$. We define $\mathcal{X}_{k,b} = \{X_{k,b} : b \in B\}$ as the indices of the symbols in $X$ that belong to $X_{k,b}$.

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

In BICM, the coded bits $Q = [Q_1, Q_2, \ldots, Q_m]$ 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] (31) [30] (eq. (8))

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

Using (33), we obtain the conditional probabilities

$$P_{X|Q_k}(x|b) = \begin{cases} \frac{p_i}{P_{Q_k}(b)}, & \text{if } x \in X_{k,b} \\ 0, & \text{if } x \notin X_{k,b} \end{cases} \quad (34)$$

for $k = 1, \ldots, m$ and $b \in B$. 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 $X$ are included in $X_{k,b}$.

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

$$P_{X|Q_k}(x_i|b) = \begin{cases} \frac{1}{2}, & \text{if } x_i \in X_{k,b} \\ 0, & \text{if } x_i \notin X_{k,b} \end{cases}. \quad (35)$$

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

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

$$0 \leq A_{X_{k,b}} \leq (M/2 - 1) \quad (36)$$

which differ from the corresponding bounds on $A_X$ in (13).

B. Binary Labelings for BICM

The natural binary code (NBC) [21] (Sec. II-B) is defined as the binary labeling $l$ 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 $X = \mathcal{E}$ [21]. Theorem 14]. A labeling $l$ is said to be a GC if for all $i, j$ such that $|x_i - x_j| = d$, the binary labels $q_i$ and $q_j$ 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

$$C_{X,1} \triangleq \frac{1}{M} \sum_{k=1}^{M} \sum_{b \in B} \sum_{x_i \in X_{k,b}} \sum_{w \in \mathcal{W}} A_{\mathcal{X}_{k,b}}^{(i)}(wd). \quad (37)$$

For a given subconstellation $X_{k,b}$, the two inner sums in (37) consider all the constellation points in the subconstellation $X_{k,b}$ at MED from $x_i \in X_{k,b}$. Thus, the quantity $C_{X,1}$ corresponds to twice the total number of different bits between the labels of constellation symbol pairs at MED. Using this
Moreover, $B_{P_X}^{(i)}(\delta)$ in (15) and $D_{P_{X_{k,b}}}^{(i)}(\delta)$ in (42) are related via

$$B_{P_X}^{(i)}(\delta) = D_{P_{X_{k,b}}}^{(i)}(\delta) + D_{P_{X_{k,b}}}^{(i)}(\delta).$$

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

$$D_{P_{X_{k,b}}} = C_{X,t,I}.$$

### 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 $I$, the BICM-GMI is defined as [22] (eq. (24))

$$I_{P_{X_{k,b}}}^{(1)}(\rho) \triangleq \sum_{k=1}^{m} \left( P_{X_k} - \sum_{b \in B} P_{Q_k}(b) I_{P_{X_{k,b}}}^{(1)}(\rho) \right)$$

and twice its derivative as

$$M_{P_{X_{k,b}}}^{(1)}(\rho) \triangleq 2 \frac{dI_{P_{X_{k,b}}}^{(1)}(\rho)}{d\rho}$$

In these expressions, $I_{P_{X_{k,b}}}^{(1)}(\rho)$ and $M_{P_{X_{k,b}}}^{(1)}(\rho)$ are defined, in analogy to (6)–(7), as

$$I_{P_{X_{k,b}}}^{(1)}(\rho) \triangleq E_{X,Y} \left[ \log \frac{P_{X_k}(X,Y|X) / P_{X_k}(Y)}{P_{Q_k}(b) I_{P_{X_{k,b}}}^{(1)}(\rho)} \right]$$

and

$$M_{P_{X_{k,b}}}^{(1)}(\rho) \triangleq E_{X,Y} \left[ (X - X_{P_{X_{k,b}}}^{(1)}(\rho))^2 \right],$$

where $X$ now follows the distribution $P_{X_{k,b}}$ and $Y$ is the random variable resulting from transmitting $X \in X_{k,b}$ over the AWGN channel (4). With these definitions, a relation corresponding to (18) between the MI and the MMSE holds also for $I_{P_{X_{k,b}}}^{(1)}(\rho)$ and $M_{P_{X_{k,b}}}^{(1)}(\rho)$, and so do theorems analogous to Theorems (11,12) which will be used in the proofs of Theorems (4,5).

Like the MI, the BICM-GMI also tends to $H_{P_X}$ as $\rho$ tends to infinity. The following theorem shows how fast $I_{P_{X_{k,b}}}^{(1)}(\rho)$ converges to $H_{P_X}$.

**Theorem 4:** For any $P_X$ and $I$

$$\lim_{\rho \to \infty} \frac{H_{P_X} - I_{P_{X_{k,b}}}^{(1)}(\rho)}{Q(\sqrt{\rho}d/2)} = \pi D_{P_{X_{k,b}}}.$$

where $D_{P_{X_{k,b}}}$ is given in (41).

**Proof:** The proof is given in Appendix D.

The following theorem characterizes the asymptotic behavior of $M_{P_{X_{k,b}}}^{(1)}(\rho)$.

Even though the BICM-GMI is fully determined by the bit probabilities ($P_{Q_k}(0), P_{Q_k}(1), P_{Q_m}(0)$), we express it as a function of the input distribution $P_X$ in (13).

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 $M_{P_{X_{k,b}}}^{(1)}(\rho)$.
Theorem 5: For any $P_X$ and $l$

$$\lim_{\rho \to \infty} \frac{M_{P_X,l}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{\pi d^2}{4} D_{P_X,l}$$

where $D_{P_X,l}$ is given in (51).

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

$$\lim_{\rho \to \infty} \frac{M_{P_X,l}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{\pi d^2}{4} B_{P_X} - \sum_{b \in B} P_{Q_k(b)} B_{P_{X_k,b}}$$

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

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

$$B_{P_{X,l}}(\rho) \triangleq \frac{1}{m} \sum_{k=1}^{m} \Pr \{ \hat{Q}_{k}^{\text{MAP}}(Y) \neq Q_k \}$$

(53)

where $Q_k$ is the transmitted bit and $\hat{Q}_{k}^{\text{MAP}}(Y)$ is a hard-decision on the bit, i.e., $\{\hat{Q}_{1}^{\text{MAP}}(y), \ldots, \hat{Q}_{m}^{\text{MAP}}(y)\} = \Phi^{-1}(\hat{X}^{\text{MAP}}(y))$ with $\hat{X}^{\text{MAP}}(y)$ given by (52) The next theorem characterizes the asymptotic behavior of the BEP in (53).

Theorem 6: For any $P_X$ and $l$

$$\lim_{\rho \to \infty} \frac{B_{P_{X,l}}(\rho)}{Q(\sqrt{\rho}d/2)} = \frac{D_{P_{X,l}}}{m}$$

(54)

where $D_{P_{X,l}}$ is given in (51).

Proof: The proof is given in Appendix E.

Similarly to (23)–(24), we can use Theorems 4–6 to show that, at high SNR, the BICM-GMI, twice its derivative, and the BEP behave as

$$I_{B_{X,l}}^{\text{BI}}(\rho) \approx H_{P_X} - \pi D_{P_X,l} Q\left(\frac{\sqrt{\rho}d}{2}\right)$$

$$M_{P_X,l}^{\text{BI}}(\rho) \approx \frac{\pi d^2}{4} D_{P_X,l} Q\left(\frac{\sqrt{\rho}d}{2}\right)$$

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

(55)

(56)

(57)

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

Example 4: Consider the constellation in Example 1 i.e., $X = \{\pm4, \pm2\}$, 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 $l_{GC} = [0, 1, 3, 2]$, which gives $A_X = C_X,l = 4$, and the three input distributions

$$p' = [1/4, 1/4, 1/4, 1/4]$$

$$p'' = [1/8, 3/8, 3/8, 1/8]$$

$$p''' = [16/25, 4/25, 1/25, 4/25]$$

which are induced by the bit probabilities listed in the second column of Table III. Table III also lists $H_{P_X}, D_{P_X,l}$, and $d$ (when the constellation is normalized to $E_s = 1$). 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 III). The corresponding curves for $H_{P_X} - [\text{BI}_{P_X,l}(\rho)]$ are shown in Fig. 6. These figures show how the coefficient $D_{P_X,l}$ 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 $X$ and $l$ and a uniform input distri-
We further define
\[ R_{P_{X},I} \triangleq \lim_{\rho \to \infty} K_{P_{X},I}^{1}(\rho) \]
\[ = \lim_{\rho \to \infty} K_{P_{X},I}^{M}(\rho) \]
where (65) follows from L'Hôpital's rule. Theorems 1 and 4 yield
\[ R_{P_{X},I} = \frac{D_{P_{X},I}}{B_{P_{X}}} \]
and due to (63)
\[ R_{P_{X},I} \geq 1. \]

In the rest of this section, we study \( R_{P_{X},I} \) in (66) for a uniform input distribution \( P_{X}^{0} \). With a slight abuse of notation, we will refer to \( R_{P_{X},I} \) as \( R_{X,I} \).

**Lemma 3:** For any labeling \( I \) and constellation \( X \), \( R_{X,I} \) is given by
\[ R_{X,I} = \frac{C_{X,I}}{A_{X}} \]
\[ \text{(68)} \]

**Proof:** Follows by using (43) and (17) in (66).

Based on Lemma 3 an upper bound on \( R_{X,I} \) can be obtained as follows.

**Theorem 7:** For any one-dimensional constellation and any labeling \( I \)
\[ C_{X,I} \leq \min (m A_{X}, (m - 1) A_{X} + M) \]
\[ \text{(69)} \]
and thus,
\[ R_{X,I} \leq \frac{\min (m A_{X}, (m - 1) A_{X} + M)}{A_{X}}. \]
\[ \text{(70)} \]

**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 \( A_{X} \leq M \). This case gives \( C_{X,I} \leq m A_{X} \). If there are more than \( M/2 \) pairs of constellation points at MED, i.e., \( A_{X} > M \), \( M/2 \) pairs can differ in \( m \) bits and the remaining \( (A_{X} - M)/2 \) pairs can differ in at most \( m - 1 \) bits, which gives \( C_{X,I} \leq M + (m - 1) (A_{X} - M) = (m - 1) A_{X} + M \). The expression in (70) follows from (69) and (63).

For an \( MPAM \) constellation, using (14), Lemma 3 and Theorem 7 specialize into
\[ R_{\mathcal{E},I} = \frac{C_{\mathcal{E},I}}{2(M - 1)}, \]
\[ \text{(71)} \]
\[ R_{\mathcal{E},I} \leq m - \frac{M - 2}{2M - 2}. \]
\[ \text{(72)} \]

Furthermore, if the \( MPAM \) constellation is labeled with the NBC, we obtain via (40)
\[ R_{\mathcal{E},I_{\text{NBC}}} = \frac{2M - m - 2}{M - 1}. \]
\[ \text{(73)} \]

**Example 5:** In Fig. 7 we show the functions \( K_{P_{X},I}^{1}(\rho) \) and \( K_{P_{X},I}^{M}(\rho) \) in (61) and (62), respectively, for a \( 4PAM \) constellation with a uniform input distribution \( P_{X} = P_{X}^{0} \), \( A_{X} = 6 \) and the three labelings that give different BICM-GMI: \( l_{\text{GC}} = [0,1,3,2] \), \( l_{\text{NBC}} = [0,1,2,3] \), and \( l_{\text{AGC}} = \ldots \)
The values of \( R_{E,I} \) 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 \( K_{PAM}^I(\rho) \) and \( K_{PAM}^M(\rho) \) allow us to study different labelings at high SNR. Observe that the GC (i.e., \( l_{GC} \)), gives \( R_{E,I_{GC}} = 1 \), and that \( l_{AGC} \) achieves the upper bound in (72), i.e., \( R_{E,I_{AGC}} = 5/3 \).

The function \( K_{PAM}^I(\rho) \) 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 \( M_{PAM}^I(\rho) \) as \( \rho \) tends to zero, which agrees with [20, 21] Theorem 14.\(^9\)

Recall that the best labeling in terms of \( H_{PAM}^I(\rho) \) at low SNR is by (18) the worst one in terms of \( M_{PAM}^I(\rho) \) (i.e., the one that maximizes \( M_{PAM}^I(\rho) \)). Furthermore, a labeling that gives a high \( M_{PAM}^I(\rho) \) at low SNR tends to yield a low \( M_{PAM}^I(\rho) \) at high SNR, since

\[
\int_0^\infty M_{PAM}^I(\rho)d\rho = 2\log M
\]

is constant for a given constellation.

Example 6: In Fig. 8 we show the function \( M_{PAM}^I(\rho) \) for 8PAM (\( P_X = P_X^{16}, A_X = 14 \)) and all the 458 labelings that give a different BICM-GMI [23]. In this figure, 12 possible values of \( R_{E,I} \) in (71) are clearly visible, which coincide with the results in [23] Fig. 3.\(^9\) Using \( M_{PAM}^I(\rho) \), the 12 values of \( R_{E,I} \) in Fig. 8 also translate into 12 different asymptotic BEP curves, which were recently reported in [33] Fig. 4. The value \( R_{E,I_{NBC}} \) obtained using (74) is also shown. A careful examination of Fig. 8 reveals that there are three labelings minimizing \( R_{E,I} \). These are the three nonequivalent GCs (in terms of BEP).\(^{28}\) Table I: the BRGC \( I = [0, 1, 3, 2, 6, 7, 5, 4] \), \( l = [0, 1, 3, 2, 6, 4, 5, 7] \), and \( I = [0, 1, 3, 7, 5, 4, 6, 2] \).

Example 7: Motivated by [21] Fig. 6, we present in Fig. 9 an approximation for the PMF \( \Pr\{R_{X,I} = r\} \) for 16PAM obtained by randomly generating \( 10^3 \) labelings. This figure shows that most of the possible labelings are not Gray. For \( M = 16 \), we obtain \( R_{E,I_{NBC}} = 26/15 \), 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 \( X \) and a uniform input distribution, a labeling \( I \) is asymptotically optimal (AO) in terms of BICM-GMI if

\[\text{BICM-GMI} \]
it satisfies $R_{X_I} = 1$. Intuitively, an AO labeling is a binary labeling for which the BICM-GMI approaches $H_{PX}$ as fast as the MI does for the same constellation $X$.

By inspection of (74), we see that the NBC for $M$PAM 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 conjuncture of the optimality of GCs at high SNR in terms of BICM-GMI [13, Sec. III-C]. (It has previously been disproved for low to medium SNRs [12].)

**Theorem 8:** For any constellation $X$ 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, (19) holds with equality, and by (25), $R_{X_I} = 1$. 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, $C_{X_I} > A_X$, and therefore, $R_{X_I} > 1$. □

**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 $M$PAM 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 $R_{X_I}$. In what follows, we show that, for $M$PAM constellations, it is always possible to construct a labeling that maximizes $R_{X_IA}$, i.e., a labeling that achieves the upper bound in (72).

We define the set of all possible values that $C_{X_I}$ can take as $\mathcal{C}_X$, where

$$|\mathcal{C}_X| \leq \frac{1}{2} \min \{ (m-1)A_X + 2, (m-2)A_X + M + 2 \}. \tag{75}$$

This inequality follows because $C_{X_I}$ 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 $X = E$, by using (14) in (75), we obtain

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

For 4PAM we have $|\mathcal{C}_E| \leq 3$ and for 8PAM we have $|\mathcal{C}_E| \leq 12$, 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 $|\mathcal{C}_E| \leq 39$. However, Fig. 8 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 $W_m$, where the $i$th row is the binary label for $x_i$, where $W_1 = [0,1]^T$, and where the following steps construct $W_m$ from $W_{m-1}$:

1. Step 1 Reverse the $M/2$ rows in $W_{m-1}$, and append them below $W_{m-1}$ to construct a new matrix $W'_m$ with $M$ rows and $m - 1$ columns.
2. Step 2 Append the length $M$ column vector $[0,1,0,1,\ldots,0,1]^T$ to the left of $W'_m$ to create $W''_m$, with $M$ rows and $m$ columns.
3. Step 3 Negate all bits in the lower half of $W''_m$ to obtain $W_m$.

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 $W_m$ are unique, and thus, the AGC is a valid labeling.

**Proof:** Consider the above construction of an AGC. Assume that $W_{m-1}$ is a valid labeling (all rows are unique) where every odd row differs in $m - 1$ bits compared to the row below, $W_1$ 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 $W'_m$ is identical to an even row in the lower half of $W_m$, which directly implies that all rows of $W'_m$ in Step 2 are unique.

Thus, $W''_m$ is a valid labeling. It also implies that every odd row of $W''_m$ differs in $m$ bits compared to the row below, since the corresponding rows of $W'_m$ in Step 2 are unique.

Inverting all the bits in the lower half of $W''_m$ is therefore equivalent to swapping every odd row in the lower half of $W''_m$ with the row below. This operation makes $W_m$ 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 $M$PAM constellations.

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

$$R_{E_{IA}} = m - M - \frac{2M - 2}{2M - 2}. \tag{77}$$

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

The labeling $L_1$ in Example 3 and Fig. 7 (i.e., $W_2$ in Fig. 10) is the AGC for 4PAM with $R_{E_{IA}} = 5/3$ given by (77). For 8PAM, the AGC is $L_{AGC} = [0,7,2,5,6,1,4,3]$ (W_3 in Fig. 10), whose corresponding function $K_{PAM_{AGC}}(\rho)$ is shown in Fig. 8, with $R_{E_{IA}} = 18/7$.

For $M = 16$, the labeling that maximizes $R_{E_{IA}}$ ($R_{E_{IA}} = 106/30 \approx 3.53$) is the AGC $W_4$ (as shown by Theorem 9), which can be constructed as described before. It can be further shown that the labeling with the second largest $R_{E_{IA}}$ ($R_{E_{IA}} = 104/30 \approx 3.47$) can be constructed by reversing the order of the three first rows of the AGC $W_4$. 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
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 asymptotic derivative, and BEP. Our results show that, as the SNR tends to infinity, all these quantities converge to their asymptotes proportionally to $Q(\sqrt{p_d}/2)$, 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 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 MPM 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]

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

It follows that

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

The MI in (6) can be expressed as

$$\text{I}_{P_X}(\rho) = \sum_{i \in \mathcal{I}_X} p_i \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-\sqrt{\rho \delta^i})^2} \cdot \log \sum_{j \in \mathcal{I}_X} p_j e^{-\frac{1}{2}(y-\sqrt{\rho \delta^j})^2} dy$$ (81)

$$= -\sum_{i \in \mathcal{I}_X} p_i \int_{-\infty}^{\infty} e^{-\frac{y^2}{\pi}} \log \sum_{\delta \in \mathcal{D}_X^{(i)}} p_j e^{-\sqrt{\rho \delta^i}} \cdot e^{-\frac{y^2}{4\rho \delta^i}} dt$$ (82)

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

$$\mathcal{D}_X^{(i)} = \{ x_i - x : x \in \mathcal{X} \}. \quad (83)$$

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

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

where

$$V_i(\rho) \triangleq \int_{-\infty}^{\infty} e^{-\frac{t^2}{\pi}} \log \sum_{\delta \in \mathcal{D}_X^{(i)}} R_{P_X}(\delta) \cdot e^{-\sqrt{\rho \delta^i} \cdot e^{-\frac{t^2}{4\rho \delta^i}}} dt$$ (85)
and
\[
R^{(i)}_{x} (\delta) \triangleq \left( \frac{B^{(i)}_x (\delta)}{p_i} \right)^2
\]
\[
= \begin{cases} 
\frac{p_i}{p}, & \text{if } \exists x_j \in X: x_i - x_j = \delta \\
0, & \text{otherwise}
\end{cases}
\]
\[
\text{(86)}
\]
\[
\text{(87)}
\]
Combining (80) and (84) yields
\[
\lim_{\rho \to \infty} \frac{H_{P_x} - l_{P_x} (\rho)}{G (\sqrt{\rho/2})} = \lim_{\rho \to \infty} \frac{H_{P_x} - l_{P_x} (\rho)}{G (\sqrt{\rho/2})} = \sum_{i \in L_x} \lim_{\rho \to \infty} \frac{G (V_i (\rho))}{G (\sqrt{\rho/2})}
\]
\[
\text{(88)}
\]
\[
\text{(89)}
\]
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 \( r = d \sqrt{\rho / \delta} \), we obtain
\[
\lim_{\rho \to \infty} \frac{G (V_i (\rho))}{G (\sqrt{\rho/2})} = 2 \left[ \lim_{r \to \infty} F^-_i (r) + \lim_{r \to \infty} F^+_i (r) \right]
\]
\[
\text{(90)}
\]
where
\[
F^-_i (r) \triangleq \int_{-\infty}^{0} re^{2r - t^2} \log \sum_{\delta \in D^{(i)}_x} R^{(i)}_x (\delta) \cdot e^{-4r + 4r^2 \frac{d^2}{\delta}} dt
\]
\[
\text{(91)}
\]
and
\[
F^+_i (r) \triangleq \int_{0}^{\infty} re^{2r - t^2} \log \sum_{\delta \in D^{(i)}_x} R^{(i)}_x (\delta) \cdot e^{-4r + 4r^2 \frac{d^2}{\delta}} dt.
\]
\[
\text{(92)}
\]
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
\[
F^-_i (r) = \int_{-\infty}^{\infty} e^{2u - \frac{u^2}{\delta}} \log \sum_{\delta \in D^{(i)}_x} R^{(i)}_x (\delta) \cdot e^{-4u + 4u^2 \frac{d^2}{\delta}} U (du)
\]
\[
\text{(93)}
\]
where
\[
U (\delta) \triangleq \delta \left( \frac{\delta}{d^2} - 1 \right).
\]
\[
\text{(94)}
\]
Note that \( U (\delta) \geq 0, \forall \delta \in D^x \). Defining
\[
f^-_i (r, u) \triangleq h (x^2 - u) \cdot e^{2u - \frac{u^2}{\delta}} \log \left( 1 + \sum_{\delta \in D^{(i)}_x} R^{(i)}_x (\delta) \cdot e^{-4u + 4u^2 \frac{d^2}{\delta}} \right)
\]
\[
\text{(95)}
\]
with \( D^x \triangleq D^{(i)}_x \setminus \{0\} \) 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 (92) can be written as
\[
F^-_i (r) = \int_{-\infty}^{\infty} f^-_i (r, u) du.
\]
\[
\text{(96)}
\]
Note that, for every \( r > 0 \), the function \( u \mapsto f^-_i (r, u) \) is nonnegative. Further note that \( U (d) = 0 \),
\[
\lim_{r \to \infty} f^-_i (r, u) = e^{2u} \log \left( 1 + R^{(i)}_x (d) \cdot e^{-4u} \right), \ u \in \mathbb{R}.
\]
\[
\text{(97)}
\]
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
\[
\lim_{r \to \infty} F^-_i (r) = \lim_{r \to \infty} \int_{-\infty}^{\infty} f^-_i (r, u) du
\]
\[
= \int_{-\infty}^{\infty} e^{2u} \log \left( 1 + R^{(i)}_x (d) \cdot e^{-4u} \right) du
\]
\[
\text{(98)}
\]
\[
\text{(99)}
\]
\[
\text{(100)}
\]
where (100) is obtained from (97) and (101) follows from the substitution \( x^2 = R^{(i)}_x (d) e^{-4u} \) 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 \( r > 0 \)
\[
0 \leq f^-_i (r, u) \leq g^-_i (u), \ u \in \mathbb{R}
\]
\[
\text{(102)}
\]
where
\[
g^-_i (u) \triangleq \begin{cases}
\begin{align*}
\text{e}^{2u} \log \left( \frac{M e^{-4u^2 / d^2}}{d^2} \right), & \text{if } u < 0 \\
\text{e}^{2u} \log \left( 1 + \frac{M - 1}{d^2} \text{e}^{-4u} \right), & \text{if } u \geq 0
\end{align*}
\end{cases}
\]
\[
\text{(103)}
\]
and \( d \) is the maximum ED of the constellation, i.e., \( d \triangleq \max_{x_i, x_j \in X} |x_i - x_j| \). Furthermore,
\[
\int_{-\infty}^{\infty} g^-_i (u) du < \infty.
\]
\[
\text{(104)}
\]
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 \( e^{-\frac{u^2}{\delta}} \leq 1 \) and \( R^{(i)}_x (d) \leq 1/p_i \) to upper-bound (79) as
\[
f^-_i (r, u) \leq h (r^2 - u) e^{2u} \log \left( 1 + \sum_{\delta \in D^{(i)}_x} \frac{e^{-4u + 4u^2 \frac{d^2}{\delta}}}{p_i} \right)
\]
\[
\leq e^{2u} \log \left( 1 + \sum_{\delta \in D^{(i)}_x} \frac{e^{-4u + 4u^2 \frac{d^2}{\delta}}}{p_i} \right)
\]
\[
\text{(105)}
\]
\[
\text{(106)}
\]
where to pass from (105) to (106) we used \( e^{-4u^2 / d^2} \leq e^{-4u} \delta \) 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
\[
f^-_i (r, u) \leq e^{2u} \log \left( 1 + \frac{M - 1}{d^2} \text{e}^{-4u} \right)
\]
\[
\text{(107)}
\]
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
\[
\int_\delta^a g_i^-(u) \, du \leq e^{2u} \log \frac{\delta}{p_i} \sum_{i \in \mathcal{I}_X} \frac{e^{-2u\delta^2/d^2}}{p_i} \int_\delta^\infty \left( \frac{M}{p_i} - \frac{\delta^2}{d^2} \right) \, du
\]
where (106) is obtained by using \( 1 < 1/p_i \) and (109) follows from \( \delta^2 < d^2, \delta \in \mathcal{D}_i \).

To prove (104), we write
\[
\int_{-\infty}^\infty g_i^-(u) \, du = \int_{-\infty}^0 g_i^-(u) \, du + \int_0^\infty g_i^-(u) \, du
\]
where from (109)
\[
\int_{-\infty}^0 g_i^-(u) \, du = \frac{1}{2} \log \frac{M}{p_i} + \frac{d^2}{d^2}
\]
and from (107)
\[
\int_0^\infty g_i^-(u) \, du \leq \int_{-\infty}^\infty e^{2u} \left( 1 + \frac{M - 1}{p_i} \right) e^{-4u} \, du
\]
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 (96) can be computed along the same lines by substituting \( t = u/r + r \) in (92), which gives
\[
\lim_{r \to \infty} F_i^+(r) = \frac{\pi \sqrt{R_i^{(i)}(d)}}{2}
\]
(113)
Combining (101) and (113) with (96) and (89) yields
\[
\lim_{\rho \to \infty} \frac{H_{P_X} - 1_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)} = \frac{1}{\pi} \sum_{i \in \mathcal{I}_X} p_i \pi \left( \sqrt{R_i^{(i)}(d)} + \sqrt{R_i^{(i)}(-d)} \right)
\]
(114)
which in view of (86) and (10) is equal to \( \pi B_{P_X} \). This proves Theorem 1.

APPENDIX B
PROOF OF THEOREM 2

For the AWGN channel in (1), the conditional mean estimator is given by
\[
\hat{X}_{ME}(y) = \sum_{i \in \mathcal{I}_X} \frac{p_i x_i e^{-y/(\sqrt{\rho \sigma_i^2})^2}}{\sum_{j \in \mathcal{I}_X} p_j e^{-y/(\sqrt{\rho \sigma_j^2})^2}}.
\]
(115)
By using (115) in (7), we obtain
\[
M_{P_X}(\rho) = \sum_{i \in \mathcal{I}_X} p_i \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-y/(\sqrt{\rho \sigma_i^2})^2} \left( \sum_{j \in \mathcal{I}_X} p_j x_i - x_j \right) e^{-y/(\sqrt{\rho \sigma_j^2})^2} \, dy
\]
(116)
where
\[
\sum_{i \in \mathcal{I}_X} p_i \hat{V}_i(\rho)
\]
(117)
and where \( R_i^{(i)}(\delta) \) is given by (88). To pass from (116) to (117) we used the substitution \( y - \sqrt{\rho \sigma_i} = \sqrt{2t} \).

Using (117), we obtain
\[
\lim_{\rho \to \infty} \frac{M_{P_X}(\rho)}{Q(\sqrt{\rho d}/2)} = \sum_{i \in \mathcal{I}_X} p_i \lim_{\rho \to \infty} \hat{V}_i(\rho) = \frac{1}{\pi} \sum_{i \in \mathcal{I}_X} p_i \pi \left( \sqrt{R_i^{(i)}(d)} + \sqrt{R_i^{(i)}(-d)} \right)
\]
(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 omit some intermediate steps.

Using (118), (79) and the substitution \( r = d\sqrt{\rho}/8 \), we have
\[
\lim_{\rho \to \infty} \frac{\hat{V}_i(\rho)}{Q(\sqrt{\rho d}/2)} = 2 \left( \lim_{r \to \infty} \hat{F}_i^{-}(r) + \lim_{r \to \infty} \hat{F}_i^{+}(r) \right)
\]
(120)
where
\[
\hat{F}_i^{-}(r) \triangleq \int_{-\infty}^{0} r e^{-r^2 - t} \left( \frac{1}{\pi} \sum_{\delta \in \mathcal{D}_i} \delta R_i^{(i)}(\delta) e^{-4rt + 4r^2 - 4\rho \sigma_i^2} \, \frac{\delta}{\rho} \right) \, dt
\]
(121)
\[
\hat{F}_i^{+}(r) \triangleq \int_{0}^{\infty} r e^{-r^2 - t} \left( \frac{1}{\pi} \sum_{\delta \in \mathcal{D}_i} \delta R_i^{(i)}(\delta) e^{-4rt + 4r^2 - 4\rho \sigma_i^2} \, \frac{\delta}{\rho} \right) \, dt
\]
(122)
We will now calculate the first limit in (120). Using the substitution \( t = u/r - r \) we express \( \hat{F}_i^{-}(r) \) in (120) as
\[
\hat{F}_i^{-}(r) = \int_{-\infty}^{0} \hat{f}_i^{-}(r, u) \, du
\]
(123)
where
\[
\hat{f}_i^{-}(r, u) \triangleq h(r^2 - u) \cdot e^{2u - \frac{\rho}{4}} \cdot \left( \frac{\sum_{\delta \in \mathcal{D}_i} \delta R_i^{(i)}(\delta) e^{-4ru + 4r^2 + 4\rho \sigma_i^2}}{1 + \sum_{\delta \in \mathcal{D}_i} R_i^{(i)}(\delta) e^{-4ru + 4r^2 + 4\rho \sigma_i^2}} \right)
\]
(124)
Recall that \( U(\delta) \) is given by (92) and \( h(x) \) is the Heavisdie’s step function. Using the fact that \( U(\delta) = 0 \) and \( U(\delta) \geq 0, \forall \delta \in \mathcal{D}_i \), we obtain
\[
\lim_{r \to \infty} \hat{f}_i^{-}(r, u) = d^2 e^{2u} \left( \frac{R_i^{(i)}(d) e^{-4u}}{1 + R_i^{(i)}(d) e^{-4u}} \right)^2
\]
(125)
As we shall prove in Lemma 5 ahead, \( u \mapsto \hat{f}_i^{-}(r, u) \) 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
\[
\lim_{r \to \infty} \tilde{F}^{-}_t(r) = \int_{-\infty}^{\infty} \lim_{r \to \infty} \tilde{f}^{-}_t(r,u) \, du = \frac{d^2 \sqrt{R^{(i)}_{P_X}(d)}}{2} \int_{0}^{\infty} \frac{x^2}{(1 + x^2)^2} \, dx = \frac{d^2 \pi \sqrt{R^{(i)}_{P_X}(d)}}{8} \tag{127}
\]
where (127) follows from (126) and the substitution \( \sqrt{R^{(i)}_{P_X}(d)} e^{-2u} = x \), and (128) follows from \( \text{eq. (3.241.5)} \).

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 (123). We obtain
\[
\lim_{r \to \infty} \tilde{F}^{+}_t(r) = \frac{d^2 \pi \sqrt{R^{(i)}_{P_X}(-d)}}{8}. \tag{129}
\]
Using (128) and (129) in (120), and combining the result with (119), (86), and (16) completes the proof.

**Lemma 6:** For any \( r > 0 \)
\[
0 \leq \tilde{f}^{-}_t(r,u) \leq \tilde{g}^{-}_t(u), \quad u \in \mathbb{R} \tag{130}
\]
where
\[
\tilde{g}^{-}_t(u) \triangleq \tilde{d} (M - 1)^2 e^{-2|u|} \tag{131}
\]
and \( \tilde{d} \) is the maximum ED of the constellation. Furthermore,
\[
\int_{-\infty}^{\infty} \tilde{g}^{-}_t(u) \, du = \frac{\tilde{d} (M - 1)^2}{p_t^2} < \infty. \tag{132}
\]

**Proof:** The first inequality in (130) follows directly from (124). To prove the second inequality in (130), we use \( e^{-\frac{2\pi^2}{7}} \leq 1, h(r^2 - u) \leq 1 \), and \( \delta \leq \tilde{d} \) to upper-bound (124) as
\[
\tilde{f}^{-}_t(r,u) \leq d^2 e^{2u} \left( \frac{\sum_{\delta \in D^T} R^{(i)}_{P_X}(\delta) e^{-4\pi^2 \delta^2 - 4r^2 U(\delta)}}{1 + \sum_{\delta \in D^T} R^{(i)}_{P_X}(\delta) e^{-4\pi^2 \delta^2 - 4r^2 U(\delta)}} \right)^2 \leq d^2 e^{2u} \left( 1 + \frac{1}{\sum_{\delta \in D^T} R^{(i)}_{P_X}(\delta) e^{-4\pi^2 \delta^2 - 4r^2 U(\delta)}} \right)^{-2} \tag{133}
\]
Since \( R^{(i)}_{P_X}(\delta) < 1/p_t \) and \( e^{-4\pi^2 \delta^2 - 4r^2 U(\delta)} \leq 1 \), we can further upper-bound (134) as
\[
\tilde{f}^{-}_t(r,u) < d^2 e^{2u} \left( 1 + \frac{p_i}{\sum_{\delta \in D^T} e^{-4\pi^2 \delta^2}} \right)^{-2} \tag{135}
\]
and
\[
< \frac{\tilde{d}^2 e^{2u}}{p_t^2} \left( 1 + \frac{1}{\sum_{\delta \in D^T} e^{-4\pi^2 \delta^2}} \right)^{-2} \tag{136}
\]
where to pass from (135) to (136) we used \( p_i < 1 \).

If \( u \geq 0 \), we have
\[
\tilde{f}^{-}_t(r,u) < \frac{d^2 e^{2u}}{p_t^2} \left( 1 + \frac{1}{(M - 1) e^{-4u}} \right)^{-2} \tag{137}
\]
\[
< \frac{d^2 e^{2u}}{p_t^2} \left( \frac{1}{(M - 1) e^{-4u}} \right)^{-2} \tag{138}
\]
\[
= \frac{d^2 (M - 1)^2}{p_t^2} e^{-6u} \tag{139}
\]
\[
< \frac{d^2 (M - 1)^2}{p_t^2} e^{-2|u|} \tag{140}
\]
where to pass from (136) to (137) all the exponentials are replaced by the one with the largest argument. If \( u \leq 0 \), (136) can be upper-bounded as
\[
\tilde{f}^{-}_t(r,u) < \frac{\tilde{d}^2 e^{2u}}{p_t^2} \tag{141}
\]
\[
\leq \frac{d^2 (M - 1)^2}{p_t^2} e^{-2|u|} \tag{142}
\]
where (141) follows from discarding the sum of exponentials in (136). Combining (140) and (142) proves (131).

**APPENDIX C**

**PROOF OF THEOREM 5**

Using Bayes’ rule, \( \hat{X}^{MAP}(y) \) in (9) can be expressed as
\[
\hat{X}^{MAP}(y) = \arg \max_{x \in \mathcal{X}} \{ f_{Y|X}(y|x) P_X(x) \} \tag{143}
\]
\[
= x_j, \quad \text{if } y \in \mathcal{Y}_j(\rho), \tag{144}
\]
where \( \mathcal{Y}_j(\rho) \) is the decision region for the symbol \( x_j \) with \( j = 1, \ldots, M \). For sufficiently large \( \rho \), these decision regions can be written as
\[
\mathcal{Y}_j(\rho) \triangleq \{ y \in \mathbb{R} : \beta_j - 1(\rho) \leq y < \beta_j(\rho) \} \tag{145}
\]
where \( \beta_i(\rho) \) with \( l = 0, \ldots, M - 1 \) are the \( M + 1 \) thresholds defining the \( M \) regions, i.e.,
\[
\beta_i(\rho) = \begin{cases} \log(p_i/p_{i+1}) + \frac{\sqrt{\pi} (x_{i+1} + x_i)}{2}, & l = 0, \ldots, M - 1 \\ +\infty, & l = M \end{cases} \tag{146}
\]
where \( \beta_i(\rho) \) for \( l = 1, \ldots, M - 1 \) in (146) is obtained by using (143) and by solving
\[
p_i f_{Y|X}(\beta_i(\rho)|x_l) = p_{l+1} f_{Y|X}(\beta_{l+1}(\rho)|x_{l+1}). \tag{147}
\]

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 \( i \in \mathcal{I}_X \)
\[
\lim_{\rho \to \infty} \frac{Q(\left|\beta_i(\rho) - \sqrt{\beta_i(\rho)}\right|)}{Q(\sqrt{\beta_i(\rho)/2})} = \begin{cases} \sqrt{R^{(i)}_{P_X}(d)}, & \text{if } l = i - 1 \\ \sqrt{R^{(i)}_{P_X}(-d)}, & \text{if } l = i \\ 0, & \text{if } l \notin \{i - 1, i\} \end{cases} \tag{148}
\]
where \( \beta_i(\rho) \) is given by (146) and \( R^{(i)}_{P_X}(\delta) \) by (87).
\textbf{Proof:} We use (146) to obtain
\[ \beta_i(\rho) - \sqrt{\beta_i x_i} = \frac{\log(p_i/p_{i+1})}{\sqrt{\beta_i(x_{i+1} - x_i)}} + \frac{\sqrt{\beta_i x_i}}{2} \] (149)

where for any \( i, l \)
\[ \epsilon_{i,l} \triangleq x_{i+1} + x_l - 2 x_i. \] (150)

Using (149) and (79), we form the ratio
\[ \frac{G(\beta_i(\rho) - \sqrt{\beta_i x_i})}{G(\sqrt{\beta_i d/2})} = \left( \frac{\rho d(x_{i+1} - x_i)}{[2\log(p_i/p_{i+1}) + \rho \epsilon_{i,l}(x_{i+1} - x_i)]} \right) \cdot \left( \frac{\epsilon_{i,l} \log(p_i/p_{i+1}) - \rho \epsilon_{i,l} d^2}{2\rho(x_{i+1} - x_i)^2} \right). \] (151)

It follows from (150) that \( |\epsilon_{i,l}| \geq x_{i+1} - x_l \geq d \) for all \( i, l \), which implies that the limit
\[ \lim_{\rho \to \infty} \frac{G(\beta_i(\rho) - \sqrt{\beta_i x_i})}{G(\sqrt{\beta_i d/2})} \] (152)
exists. We distinguish between three cases:

(i) If \( l = i \) and \( x_{i+1} - x_i = d \), then \( \epsilon_{i,l} = x_{i+1} - x_l = d \) and the limit in (152) is \( e^{-\log(p_i/p_{i+1}^2)} = \sqrt{p_{i+1}/p_i} \).

(ii) If \( l = i+1 \) and \( x_{i+1} - x_l = d \), then \( \epsilon_{i,l} = x_l - x_{i+1} = -d \) and the limit in (152) is \( \sqrt{p_l/p_{i+1}} \).

(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
\[ \lim_{\rho \to \infty} \frac{G(\beta_i(\rho) - \sqrt{\beta_i x_i})}{G(\sqrt{\beta_i d/2})} = \begin{cases} \sqrt{p_{i+1}/p_i}, & \text{if } l = i \text{ and } x_{i+1} - x_i = d \\ \sqrt{p_l/p_{i+1}}, & \text{if } l = i+1 \text{ and } x_{i+1} - x_l = -d \\ 0, & \text{otherwise} \end{cases}. \] (153)

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

Returning to the proof of Theorem 3, using (144) and (145), the SEP in (83) is expressed as
\[ S_{P_X}(\rho) = \sum_{i \in X} p_i \Pr\{Y \notin Y_i(\rho) | X = x_i\} \]
\[ = \sum_{i \in X} p_i (Q(\beta_i(\rho) - \sqrt{\beta_i x_i}) + Q(\sqrt{\beta_i x_i} - \beta_{i-1}(\rho))) \] (154)
\[ \] (155)

which gives
\[ \lim_{\rho \to \infty} Q(\sqrt{\beta_i d/2}) \]
\[ = \sum_{i \in X} p_i \left( \lim_{\rho \to \infty} Q\left(\frac{\beta_i(\rho) - \sqrt{\beta_i x_i}}{\sqrt{\beta_i d/2}}\right) + \frac{Q\left(\sqrt{\beta_i x_i} - \beta_{i-1}(\rho)\right)}{Q(\sqrt{\beta_i d/2})} \right) \]
\[ = \sum_{i \in X} p_i \left( X_{P_X}(-d) + \sqrt{R_P(d)} \right) \] (156)
\[ = \sum_{i \in X} p_i \left( R_{P_X}(d) \right) \] (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 \( \rho \). The proof of Theorem 3 is completed by using (80) in (157) together with (16).

**APPENDIX D \textbf{PROOF OF THEOREM 4}\**

The following lemma will be used in this proof as well as in the proof of Theorem 3.

\textbf{Lemma 8:}
\[ \sum_{k=1}^{m} \left( B_{P_X} - \sum_{b \in B} P_{Q_k}(b) B_{P_{X,b}} \right) = D_{P_X,t} \] (158)
where \( D_{P_X,t} \) is given by (41), \( B_{P_X} \) is given by (16).

\[ B_{P_{X,b}} = \sum_{i \in I_{X,b}} \sum_{w \in W} B_{P_{X,b}}(wd) \] (159)
and
\[ B_{P_{X,b}}(\delta) = \begin{cases} \sqrt{P_{X,b}(x_j)P_{X,b}(x_i)}, & \text{if } \exists x_j \in X_{k,b}, x_i - x_j = \delta \\ 0, & \text{otherwise} \end{cases} \] (160)
with \( P_{X,b}(x) \) given by (34).

\textbf{Proof:} From (34), \( p_i = P_{Q_k}(b) P_{X,b}(x_i) \) for any \( b \in B \), \( k = 1, \ldots, m \) and \( i \in I_{X,b} \). Hence, using (160) and (42) gives
\[ D_{P_X,b}(\delta) = P_{Q_k}(b)B_{P_{X,b}}(\delta). \] (161)

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

\[ \sum_{k=1}^{m} \left( B_{P_X} - \sum_{b \in B} P_{Q_k}(b) B_{P_{X,b}} \right) = \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X,b}} \sum_{w \in W} (B_{P_{X,b}}(wd) - D_{P_{X,b}}(wd) \right) \] (162)
\[ = \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X,b}} \sum_{w \in W} D_{P_{X,b}}(wd) \] (163)
where (163) is obtained using (43). The proof is completed by comparing (163) with (41). \( \square \)

Using the expression for the BICM-GMI (45), we have
\[ H_{P_X} = \frac{1}{P_{X,T}}(\rho) \]
\[ = \sum_{k=1}^{m} (H_{P_X} - 1_{P_X}(\rho)) \]
\[ - \sum_{k=1}^{m} \sum_{b \in B} P_{Q_k}(b)(H_{P_{X,b}} - 1_{P_{X,b}}(\rho)) \]
\[ - (m - 1)H_{P_X} + \sum_{k=1}^{m} \sum_{b \in B} P_{Q_k}(b)H_{P_{X,b}}. \] (164)
The last term on the r.h.s. of (164) is zero because
\[
\sum_{k=1}^{m} \sum_{b \in B} P_{Q_k}(b) H_{P_{X_{k,b}}} = 0
\]
(165)

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 \(Q(\sqrt{p}d/2)\) 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
\[
\lim_{\rho \to \infty} \frac{H_{P_X} - \frac{1}{m} \sum_{j=1}^{m} B_{P_X = j}}{Q(\sqrt{p}d/2)} = \pi \sum_{k=1}^{m} \left( B_{P_X} - \sum_{b \in B} P_{Q_k}(b) B_{P_{X_{k,b}}} \right)
\]
(170)

which in view of Lemma [8] completes the proof.

**APPENDIX E**

**PROOF OF THEOREM 6**

The BEP in (53) is expressed as
\[
B_{P_X,j}(\rho) = \frac{1}{m} \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X_k,b}} p_i \Pr\{Q_k^{MAP} \neq q_{k,i} | X = x_i\}
\]
(171)

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 \(b\) at bit position \(k\), by using the fact that \(Y_j(\rho)\) are disjoint, and \(\cup_{j \in I_X} Y_j = \mathbb{R}\). Furthermore, we express (172) in terms of pairwise error probabilities (PEP) as
\[
B_{P_X,j}(\rho) = \frac{1}{m} \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X_k,b}} p_i \Pr\{Y_j(\rho) | X = x_i\}
\]
(173)

where
\[
\Pr\{Y_j(\rho) | X = x_i\} = \frac{1}{m} \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X_k,b}} p_i \Pr\{Y_j(\rho) | X = x_i\}
\]
(174)

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
\[
\lim_{\rho \to \infty} \frac{H_{P_X} - \frac{1}{m} \sum_{j=1}^{m} B_{P_X = j}}{Q(\sqrt{p}d/2)} = \pi \sum_{k=1}^{m} \left( B_{P_X} - \sum_{b \in B} P_{Q_k}(b) B_{P_{X_{k,b}}} \right)
\]
(170)

which in view of Lemma [8] completes the proof.

\[
S_4 = \sum_{j \in I_X, j < i} \lim_{\rho \to \infty} \frac{Q(\sqrt{p}x_j - \beta_j(\rho))}{Q(\sqrt{p}d/2)}
\]
(179)

where all the arguments of the Q-functions in (179) are positive for large enough \(\rho\).

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 \(Q(\sqrt{p}x_j - \beta_j(\rho)) = Q(|\beta_j(\rho) - \sqrt{p}x_j|)\) and Lemma [7] to obtain \(S_2 = 0\). Similarly, using Lemma [7], we conclude that the only nonzero contribution to \(S_1\) can come from the term
Combining these results and using the counting function \( \left( 11 \right) \), we express \( \left( 178 \right) \) as

\[
\lim_{\rho \to \infty} \frac{B_{FX,\rho}(\rho)}{Q(\sqrt{\rho d/2})}
= \frac{1}{m} \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X,k,b}} p_l \left( A_{X_k}(i) Q(\beta_1(\rho) Q(\sqrt{\rho d/2})) \left( \lim_{\rho \to \infty} \frac{Q(\beta_1(\rho) Q(\sqrt{\rho d/2}))}{Q(\sqrt{\rho d/2})} \right) \right)
+ A_{X_k}(i) Q\left( \lim_{\rho \to \infty} \frac{Q(\beta_1(\rho) Q(\sqrt{\rho d/2}))}{Q(\sqrt{\rho d/2})} \right)
= \frac{1}{m} \sum_{k=1}^{m} \sum_{b \in B} \sum_{i \in I_{X,k,b}} p_l \left( A_{X_k}(i) \frac{Q(\beta_1(\rho) Q(\sqrt{\rho d/2}))}{Q(\sqrt{\rho d/2})} \right)
+ A_{X_k}(i) \frac{Q(\beta_1(\rho) Q(\sqrt{\rho d/2}))}{Q(\sqrt{\rho d/2})}
\]  

where to pass from (183) to (184) we used Lemma 7. Furthermore, by combining (87) and (52) we obtain

\[
p_l A_{X_k}(i) \frac{Q(\beta_1(\rho) Q(\sqrt{\rho d/2}))}{Q(\sqrt{\rho d/2})} = D_{FX,\rho}(\delta)
\]  

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.