High SNR Bounds for the BICM Capacity

Alex Alvarado, Fredrik Brännström[§], and Erik Agrell[§]
Department of Engineering, University of Cambridge, UK
[§]Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden
alex.alvarado@ieee.org, {*fredrik.brannstrom,agrell*}*@chalmers.se*

***Abstract*—In this paper, different aspects of the bit-interleaved coded modulation (BICM) capacity for the Gaussian channel are analyzed. Analytical bounds for the BICM capacity are developed. These bounds suggest that the BICM capacity at high signal-to-noise ration (SNR) is determined by the multiplicity of the minimum Euclidean distance over all the subconstellations generated by the mapper. Based on this observation, we conjecture that for any constellation, the highest BICM capacity at high SNR is always obtained by a Gray code, if one exists. Ready-to-use expressions based on Gauss–Hermite quadratures to compute the coded modulation and BICM capacities for any SNR are also presented. Using these expressions, it is shown that the BICM capacity is in general a nonconvex, nonconcave function of the input bit distribution. For 8PAM and 8PSK, there exist 12 and 7 classes of mappings, respectively, with equivalent high-SNR behavior, of which the best class comprises all Gray codes.**

I. Introduction

The first breakthroughs for coding in the bandwidth-limited regime came with Ungerboeck's trellis-coded modulation (TCM) [1] and Imai and Hirakawa's multilevel coding (MLC) [2]. The next breakthrough came in 1992, when Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [3]–[5]. BICM is usually referred to as a pragmatic approach for coded modulation (CM) design and it is used in almost all of the current wireless communications standards, e.g., HSPA, IEEE 802.11a/g/n, and the DVB standards (DVB-T2/S2/C2).

For a given signal-to-noise ratio (SNR), the maximum rate of CM systems is defined in term of the average mutual information (AMI) of a discrete-input continuous output (DICO) channel. Closed-form expressions for the AMI are in general unknown, which necessitates numerical computation methods. Such computations involve the evaluation of multidimensional expectations with unbounded supports. Generic algorithms for multidimensional integration, such as Riemann sums and Monte Carlo integration, can be applied. In this paper, a third integration method is applied, tailored for the specific integrand in AMI expressions, using Gauss–Hermite (GH) quadratures. The main advantage of this method is that it offers the best complexity/accuracy tradeoff.

The use of GH quadratures to compute the capacities in general, and CM or BICM capacities in particular, is a well known method used in the literature. For example, they have been used to compute the capacity of distributed antenna systems [6], the ergodic capacity of cooperative spatial multiplexing systems [7], and the CM capacity for N-dimensional constellations and uniform input distributions [8], [9]. Recently, rapidly-converging series representation for the computation of the AMI for one-dimensional constellations have been investigated in [10]. For BICM, the use of GH quadratures is mentioned for example in [5, Sec. 3.4], [11, Sec. III], [12, Sec. III]; however, to the best of our knowledge, there are no explicit ready-to-use expressions available in the literature. In this paper, we fill this gap by presenting ready-to-use expressions for both CM and BICM capacities. These expressions are generalizations of the ones in [8], [9] to the BICM capacity and arbitrary input distributions.

The BICM capacity depends heavily on the mapping. The optimality of a Gray code was conjectured in [4, Sec. III-C], which was later disproved in [13]. It is shown in [13] (see also [14, Ch. 3]) that for low and medium SNR, there exist other mappings that give a higher BICM capacity. Moreover, in [13, Sec. III-C] it is conjectured that among all the Gray codes, the binary reflected Gray code (BRGC) [15] is the one that maximizes the BICM capacity. The BICM capacity for asymptotically low rates was studied in [16], where it is shown that BICM with the BRGC does not achieve the Shannon limit (SL) -1.59 dB. It was later shown in [17] that for MPAM and MQAM constellations, the natural binary code makes BICM achieve the SL. These results were generalized in [18], where general signal sets and mappings were studied.

In this paper, we develop bounds for the CM and BICM capacities, which are shown to capture their high-SNR behavior, i.e., to show that mappings can be classified into a small number of classes at high SNR. Based on the developed bounds and numerical results, we conjecture that for any constellation and high SNR, a Gray code—if it exists—is the capacity-maximizing mapping for the BICM capacity.

II. Preliminaries

A. System Model and Notation Convention

We use boldface letters $\boldsymbol{x}$ to denote row vectors $\boldsymbol{x}_i = [x_{i,1}, \ldots, x_{i,N}]$. Sets are denoted using calligraphic letters $\mathcal{C}$ and the binary set is defined as $\mathcal{B} \triangleq \{0, 1\}$. All the logarithms used are natural logarithms. Random variables are denoted by capital letters Y, the probability mass function (PMF) of the random vector $\boldsymbol{Y}$ by $P_{\boldsymbol{Y}}(\boldsymbol{y})$, and the probability density function (PDF) of the random vector $\boldsymbol{Y}$ by $p_{\boldsymbol{Y}}(\boldsymbol{y})$. The joint PDF of the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ is denoted by

Research supported by The British Academy and The Royal Society (via the Newton International Fellowship scheme), UK, and by the Swedish Research Council, Sweden (under grant #621-2006-4872).

Table I
SIX FUNCTIONS TO BE USED IN (10) TO EVALUATE THE CM AND BICM CAPACITIES.

Function	Expression	Equations
$g_{\boldsymbol{p}}^{\mathrm{CM}}(\boldsymbol{t})$	$-\pi^{-N/2}\sum_{i\in\mathcal{I}}P_{\boldsymbol{X}}(\boldsymbol{x}_i)\log\sum_{j\in\mathcal{I}}P_{\boldsymbol{X}}(\boldsymbol{x}_j)\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle-\rho\|\boldsymbol{d}_{i,j}\|^2}$	(5), (11)
$g_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})$	$-M^{-1}\pi^{-N/2}\sum_{i\in\mathcal{I}}\log\sum_{j\in\mathcal{I}}\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle-\rho\|\boldsymbol{d}_{i,j}\|^2}$	(6), (12)
$\tilde{g}_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})$	$-M^{-1}\pi^{-N/2}\sum_{i\in\mathcal{I}}\log\big(1+\mathrm{e}^{-\rho\hat{d}_{\mathcal{X}}^2}\sum_{j\in\hat{\mathcal{I}}^{(i)}}\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle}\big)$	(19)
$g_{\boldsymbol{p},\phi}^{\mathrm{BI}}(\boldsymbol{t})$	$mg_{\boldsymbol{p}}^{\mathrm{CM}}(\boldsymbol{t})+\pi^{-N/2}\sum_{k=1}^{m}\sum_{u\in\mathcal{B}}\sum_{i\in\mathcal{I}_{k,u}}P_{\boldsymbol{X}}(\boldsymbol{x}_i)\log\sum_{j\in\mathcal{I}_{k,u}}\frac{P_{\boldsymbol{X}}(\boldsymbol{x}_j)}{P_{C_k}(u)}\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle-\rho\|\boldsymbol{d}_{i,j}\|^2}$	(7), (13)
$g_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})$	$mg_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})+M^{-1}\pi^{-N/2}\sum_{k=1}^{m}\sum_{u\in\mathcal{B}}\sum_{i\in\mathcal{I}_{k,u}}\log\sum_{j\in\mathcal{I}_{k,u}}\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle-\rho\|\boldsymbol{d}_{i,j}\|^2}$	(8), (14)
$\tilde{g}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})$	$mg_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})+M^{-1}\pi^{-N/2}\sum_{k=1}^{m}\sum_{u\in\mathcal{B}}\sum_{i\in\mathcal{I}_{k,u}}\log\Big(1+\mathrm{e}^{-\rho\hat{d}_{\mathcal{X}}^2}\sum_{j\in\hat{\mathcal{I}}_{k,u}^{(i)}}\mathrm{e}^{-2\sqrt{\rho}\,\langle\boldsymbol{t},\boldsymbol{d}_{i,j}\rangle}\Big)$	(20)

$p_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x},\boldsymbol{y})$, and the conditional PDF of $\boldsymbol{Y}$ conditioned on $\boldsymbol{X}=\boldsymbol{x}$ is denoted by $p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})$. The expectation of an arbitrary function $f(\boldsymbol{X},\boldsymbol{Y})$ over the joint PDF of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is denoted by $\mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}[f(\boldsymbol{X},\boldsymbol{Y})]$.

We use $\boldsymbol{x}_i\in\mathcal{X}$ to denote the transmitted symbols, where $\mathcal{X}$ is the constellation used for transmission, $|\mathcal{X}|=M=2^m$. We define $\mathcal{I}\triangleq\{1,\dots,M\}$ which enumerates all the constellation symbols in $\mathcal{X}$, and we also define the difference between two constellation symbols as $\boldsymbol{d}_{i,j}\triangleq(\boldsymbol{x}_i-\boldsymbol{x}_j)$. The minimum distance of the constellation is denoted by $d_{\mathcal{X}}$, i.e., $\hat{d}_{\mathcal{X}}\triangleq\min_{i,j\in\mathcal{I},i\neq j}\|\boldsymbol{d}_{i,j}\|$. The input distribution is denoted by the vector $\boldsymbol{p}=[P_{\boldsymbol{X}}(\boldsymbol{x}_1),\dots,P_{\boldsymbol{X}}(\boldsymbol{x}_M)]$ and the uniform distribution is denoted by $\boldsymbol{u}=[1/M,\dots,1/M]$.

We consider a BICM system where the bits are mapped to constellation symbols using a mapper $\phi:\mathcal{B}^m\to\mathcal{X}$. The mapper is defined via the binary labelings of the symbols $\boldsymbol{x}_i$, denoted by $\boldsymbol{c}_i=[c_{i,1},\dots,c_{i,m}]\in\mathcal{B}^m$ with $i\in\mathcal{I}$. The mapping defines $2m$ *subconstellations* $\mathcal{X}_{k,u}$ for $k=1,\dots,m$ and $u\in\mathcal{B}$, i.e., $\mathcal{X}_{k,u}\triangleq\{\boldsymbol{x}_i\in\mathcal{X}:c_{i,k}=u\}$. We define $\mathcal{I}_{k,u}\subset\{1,\dots,M\}$ as the indices of the symbols in $\mathcal{X}_{k,u}$.

Assuming that the bits at the input of the modulator are independent, the input symbol probabilities are given by $P_{\boldsymbol{X}}(\boldsymbol{x}_i)=\prod_{k=1}^{m}P_{C_k}(c_{i,k})$, where C_k is the random variable representing the kth bit in the codeword mapped to $\boldsymbol{x}_i$. Consequently, the conditional input symbol probabilities, conditioned on the kth bit being u, are

$$P_{\boldsymbol{X}|C_k}(\boldsymbol{x}_i|u)=\begin{cases}\dfrac{P_{\boldsymbol{X}}(\boldsymbol{x}_i)}{P_{C_k}(u)}, & \text{if } i\in\mathcal{I}_{k,u}\\ 0, & \text{if } i\notin\mathcal{I}_{k,u}\end{cases}. \tag{1}$$

Throughout this paper, we consider a discrete-time real-valued N-dimensional additive white Gaussian noise (AWGN) channel so that each received symbol is $\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{Z}$, where $\boldsymbol{Y}\in\mathbb{R}^N$, $\boldsymbol{X}\in\mathcal{X}$, and $\boldsymbol{Z}$ is a vector of i.i.d. Gaussian random variables with zero mean and variance $N_0/2$ per dimension. We assume that the constellation $\mathcal{X}$ is normalized to unit energy, i.e, $E_{\mathrm{s}}=\mathbb{E}_{\boldsymbol{X}}[\|\boldsymbol{X}\|^2]=\sum_{i\in\mathcal{I}}P_{\boldsymbol{X}}(\boldsymbol{x}_i)\|\boldsymbol{x}_i\|^2=1$. The conditional transition PDF of the AWGN channel is

$$p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x})=(\rho/\pi)^{\frac{N}{2}}\,\mathrm{e}^{-\rho\|\boldsymbol{y}-\boldsymbol{x}\|^2}, \tag{2}$$

where the signal-to-noise ratio (SNR) is $\rho\triangleq E_{\mathrm{s}}/N_0=1/N_0$.

An MPAM constellation is defined as $\mathcal{X}_{\mathrm{PAM}}\triangleq\{\pm(M-1)\Delta,\pm(M-3)\Delta,\dots,\pm1\}$ with $\Delta^2=3/(M^2-1)$ so that $E_{\mathrm{s}}=1$. The MPSK constellation is defined as $\mathcal{X}_{\mathrm{PSK}}\triangleq\{[\cos(2\pi i/M),\sin(2\pi i/M)]:i=1,\dots,M\}$.

The AMI of a DICO memoryless channel is given by

$$I_{\boldsymbol{X};\boldsymbol{Y}}(\rho)=\mathbb{E}_{\boldsymbol{X},\boldsymbol{Y}}\left[\log\frac{p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{Y}|\boldsymbol{X})}{p_{\boldsymbol{Y}}(\boldsymbol{Y})}\right]. \tag{3}$$

The so-called "CM capacity" for a given $\mathcal{X}$ and $\boldsymbol{p}$ is denoted by $\mathsf{I}_{\boldsymbol{p}}^{\mathrm{CM}}(\rho)$ and is defined as the AMI of the DICO memoryless channel for a given constellation. Therefore, it simply corresponds to the AMI in (3), i.e., $\mathsf{I}_{\boldsymbol{p}}^{\mathrm{CM}}(\rho)\triangleq I_{\boldsymbol{X};\boldsymbol{Y}}(\rho)$.

The BICM capacity is defined for a given mapper ϕ and an input distribution $\boldsymbol{p}$ as [18, eq. (32)]

$$\begin{aligned}\mathsf{I}_{\boldsymbol{p},\phi}^{\mathrm{BI}}(\rho)&\triangleq\sum_{k=1}^{m}I_{C_k;\boldsymbol{Y}}(\rho)=\sum_{k=1}^{m}\mathbb{E}_{C_k,\boldsymbol{Y}}\left[\log\frac{p_{\boldsymbol{Y}|C_k}(\boldsymbol{Y}|C_k)}{p_{\boldsymbol{Y}}(\boldsymbol{Y})}\right]\\&=\sum_{k=1}^{m}\sum_{u\in\mathcal{B}}\sum_{i\in\mathcal{I}_{k,u}}P_{\boldsymbol{X}}(\boldsymbol{x}_i)\int_{\mathbb{R}^N}p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_i)\cdot\\&\qquad\log\frac{\sum_{j\in\mathcal{I}_{k,u}}P_{\boldsymbol{X}|C_k}(\boldsymbol{x}_j|u)p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_j)}{\sum_{j\in\mathcal{I}}P_{\boldsymbol{X}}(\boldsymbol{x}_j)p_{\boldsymbol{Y}|\boldsymbol{X}}(\boldsymbol{y}|\boldsymbol{x}_j)}\,\mathrm{d}\boldsymbol{y}.\end{aligned} \tag{4}$$

If all the bits at the input of the modulator are equally likely, i.e., $P_{C_k}(u)=1/2$ for $k=1,\dots,m$ and $u\in\mathcal{B}$, and thus, $P_{\boldsymbol{X}}(\boldsymbol{x})=1/M$, which gives the BICM capacity traditionally found in the literature, cf. [4, eq. (15)], [5, Sec. 3.2.1].

III. CAPACITIES USING GAUSS–HERMITE QUADRATURES

In this section, we show how to efficiently calculate the CM and BICM capacities using GH quadratures. We use $\boldsymbol{p}$ and $\boldsymbol{u}$ as indexes to denote an arbitrary and the uniform input distributions, respectively.

A. Simplified Expressions

Theorem 1: The CM capacity can be expressed as

$$\mathsf{I}_{\boldsymbol{p}}^{\mathrm{CM}}(\rho)=\int_{\mathbb{R}^N}\mathrm{e}^{-\|\boldsymbol{t}\|^2}g_{\boldsymbol{p}}^{\mathrm{CM}}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}, \tag{5}$$

$$\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho)=\log(M)+\int_{\mathbb{R}^N}\mathrm{e}^{-\|\boldsymbol{t}\|^2}g_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}, \tag{6}$$

where $g_{\boldsymbol{p}}^{\mathrm{CM}}(\boldsymbol{t})$ and $g_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})$ are shown in Table I.

Proof: Using (2) in (3), by splitting the logarithm of the quotient as a difference of logarithms, and by using the substitution $\boldsymbol{t}=\sqrt{\rho}(\boldsymbol{y}-\boldsymbol{x}_i)$. ∎

Theorem 2: The BICM Capacity can be expressed as

$$\mathsf{I}^{\mathrm{BI}}_{\boldsymbol{p},\phi}(\rho) = \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} g^{\mathrm{BI}}_{\boldsymbol{p},\phi}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}, \tag{7}$$

$$\mathsf{I}^{\mathrm{BI}}_{\boldsymbol{u},\phi}(\rho) = \log(M) + \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} g^{\mathrm{BI}}_{\boldsymbol{u},\phi}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}, \tag{8}$$

where $g^{\mathrm{BI}}_{\boldsymbol{p},\phi}(\boldsymbol{t})$ and $g^{\mathrm{BI}}_{\boldsymbol{u},\phi}(\boldsymbol{t})$ are shown in Table I.

Proof: Using (2) in (4) and $\boldsymbol{t} = \sqrt{\rho}(\boldsymbol{y} - \boldsymbol{x}_i)$. ■

B. The Gauss–Hermite Quadratures

For any function $g(t)$ with bounded $(2J)$th derivative, the GH quadratures [19, Sec. 7.3.4] are

$$\int_{-\infty}^{\infty} \mathrm{e}^{-t^2} g(t)\,\mathrm{d}t = \lim_{J\to\infty} \sum_{k=1}^{J} \alpha_k g(\xi_k), \tag{9}$$

where ξ_k is the kth root of $H_J(x)$, and

$$H_J(x) = J! \sum_{r=0}^{\lfloor J/2 \rfloor} \frac{(-1)^r}{r!(J-2r)!}(2x)^{J-2r},\ \alpha_k = \frac{2^{J-1}J!\sqrt{\pi}}{[JH_{J-1}(\xi_k)]^2}.$$

Tables with α_k and ξ_k for different values of J can be found for example in [19, Appendix 7.3(b)], where J is a parameter that adjusts the tradeoff between computation speed and precision. The expression in (9) can be generalized to an N-dimensional vector $\boldsymbol{t} = [t_1, \ldots, t_N]$ as[1]

$$\int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} g(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t} \approx \sum_{k_1=1}^{J} \cdots \sum_{k_N=1}^{J} g(\boldsymbol{\xi}) \prod_{n=1}^{N} \alpha_{k_n}, \tag{10}$$

where $\boldsymbol{\xi} = [\xi_{k_1}, \ldots, \xi_{k_N}]$ and the approximation in (10) is because in practice we use a finite J. All the results presented in this paper were obtained with $J = 10$, which we found to be a good tradeoff between complexity and precision.

Using (10) and Theorems 1 and 2, we obtain

$$\mathsf{I}^{\mathrm{CM}}_{\boldsymbol{p}}(\rho) \approx \sum_{k_1=1}^{J} \cdots \sum_{k_N=1}^{J} g^{\mathrm{CM}}_{\boldsymbol{p}}(\boldsymbol{\xi}) \prod_{n=1}^{N} \alpha_{k_n}, \tag{11}$$

$$\mathsf{I}^{\mathrm{CM}}_{\boldsymbol{u}}(\rho) \approx \log(M) + \sum_{k_1=1}^{J} \cdots \sum_{k_N=1}^{J} g^{\mathrm{CM}}_{\boldsymbol{u}}(\boldsymbol{\xi}) \prod_{n=1}^{N} \alpha_{k_n}, \tag{12}$$

$$\mathsf{I}^{\mathrm{BI}}_{\boldsymbol{p},\phi}(\rho) \approx \sum_{k_1=1}^{J} \cdots \sum_{k_N=1}^{J} g^{\mathrm{BI}}_{\boldsymbol{p},\phi}(\boldsymbol{\xi}) \prod_{n=1}^{N} \alpha_{k_n}, \tag{13}$$

$$\mathsf{I}^{\mathrm{BI}}_{\boldsymbol{u},\phi}(\rho) \approx \log(M) + \sum_{k_1=1}^{J} \cdots \sum_{k_N=1}^{J} g^{\mathrm{BI}}_{\boldsymbol{u},\phi}(\boldsymbol{\xi}) \prod_{n=1}^{N} \alpha_{k_n}. \tag{14}$$

C. Numerical Results

In Fig. 1, we show the BICM capacity in (13) for 8PAM and the BRGC as a function of the input bit probabilities. In order to have a symmetric input distribution $\boldsymbol{p}$ with respect to zero, and because of the structure of the BRGC, we set $P_{C_1}(u) = 1/2$, and thus, we plot the BICM capacity as a function of the other two variables. This figure shows

[1]This is the simplest (but not unique) way to do this generalization.

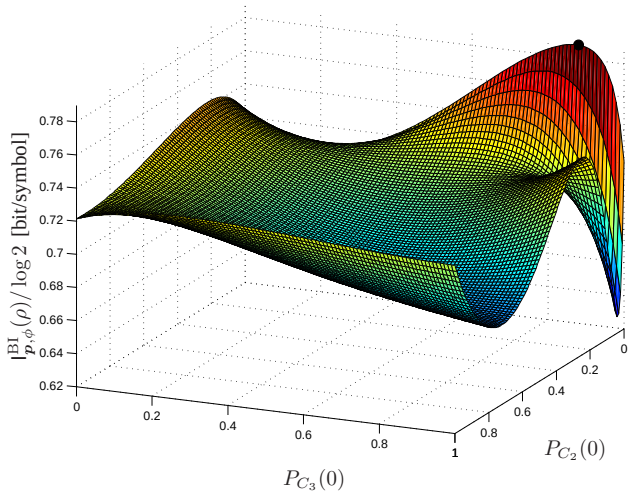


Figure 1. The BICM capacity for 8PAM with the BRGC as a function of the bit probabilities for $k = 2, 3$ and $\rho = 0$ dB. The filled black circle is $\boldsymbol{p}^*$.

that the BICM capacity is in general a nonconcave nonconvex function on the input distribution of the bits. It also shows that the optimal input distribution $\boldsymbol{p}^*$ for $\rho = 0$ dB and $P_{C_1}(0) = 1/2$ is $[P_{C_2}(0), P_{C_3}(0)] = [0, 0.88]$, i.e, $\boldsymbol{p}^* = [0, 0, 0.06, 0.44, 0.44, 0.06, 0, 0]$, which translates into a nonequally likely 4PAM constellation.

In Fig. 2 (a), we present the numerical evaluation of the BICM capacity in (14) for 8PSK, and all the possible different mappings (colored lines).[2] This figure shows that, among the $8! = 40320$ different mappings[3], there are only 26 ones that give different BICM capacity for asymptotically low rates, as previously shown in [18, Sec. V-C]. For rates above zero and below 3 bit/symbol, there are only 49 different capacities, as shown in [20, Sec. IV]. In Fig. 2 (b), we show similar results for 8PAM (only for high SNR values). For 8PAM, an exhaustive search revealed that there are 458 mappings that give different BICM capacity (colored lines).

More importantly, Fig. 2 shows that for high SNR, there is only a very limited number of *classes* of mappings with asymptotically equivalent behavior, i.e., for high SNR, all the mappings merge into a few classes (7 for 8PSK and 12 for 8PAM). The 7 classes for 8PSK can in fact be observed in [21, Table A.1]. In the following section, we give a formal explanation for these classes.

IV. Asymptotic Analysis

In this section we analyze the behavior of the CM and BICM capacity in the high SNR regime. From now on, we assume a uniform input distribution $\boldsymbol{p} = \boldsymbol{u}$. We use $\hat{\mathcal{I}}^{(i)} \triangleq \{j \in \mathcal{I} : \|\boldsymbol{d}_{i,j}\| = \hat{d}_{\mathcal{X}}\}$ and $A^{(i)}_{\mathcal{X}} = |\hat{\mathcal{I}}^{(i)}|$ to denote the number of symbols at minimum distance from the symbol $\boldsymbol{x}_i$. Also, $A_{\mathcal{X}} \triangleq \sum_{i\in\mathcal{I}} A^{(i)}_{\mathcal{X}}$ is twice the number of pairs of symbols in $\mathcal{X}$ at minimum distance. For MPAM, $A^{(i)}_{\mathcal{X}} = 1$

[2]For a clearer presentation, the capacity curves are plotted as a function of E_{b}/N_0, where E_{b} is the average bit energy, cf. [18, Sec. III].

[3]For MPSK it is enough to consider $(M-1)!$, since a rotation does not influence the AMI. However, in general there are $M!$ possible mappings.

for the end constellation symbols and $A_{\mathcal{X}}^{(i)} = 2$ for the other $M-2$, and $A_{\mathcal{X}} = 2(M-1)$. For MPSK, $A_{\mathcal{X}}^{(i)} = 2$ for $i = 1, 2, \ldots, M$, and $A_{\mathcal{X}} = 2M$. For a given mapper ϕ, we use A_ϕ to denote twice the sum of the number of pairs of symbols at minimum distance in the subconstellations $\mathcal{X}_{k,u}$, i.e., $A_\phi \triangleq \sum_{k=1}^{m} \sum_{u \in \mathcal{B}} A_{\mathcal{X}_{k,u}}$.

Lemma 3 (Lower Bounds): For any $\boldsymbol{t} \in \mathbb{R}^N$,

$$\sum_{i \in \mathcal{I}} \log \sum_{j \in \mathcal{I}} \mathrm{e}^{-\rho \|\boldsymbol{d}_{i,j}\|^2 - 2\sqrt{\rho}\,\langle \boldsymbol{t}, \boldsymbol{d}_{i,j} \rangle} \tag{15}$$

$$\geq \sum_{i \in \mathcal{I}} \log \left(1 + \mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2} \sum_{j \in \hat{\mathcal{I}}^{(i)}} \mathrm{e}^{-2\sqrt{\rho}\,\langle \boldsymbol{t}, \boldsymbol{d}_{i,j} \rangle} \right) \tag{16}$$

$$\geq \sum_{i \in \mathcal{I}} \log \left(1 + A_{\mathcal{X}}^{(i)} \mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2 - 2\sqrt{\rho} \|\boldsymbol{t}\| \hat{d}_{\mathcal{X}}} \right) \tag{17}$$

$$\geq \log \left(1 + A_{\mathcal{X}} \mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2 - 2\sqrt{\rho} \|\boldsymbol{t}\| \hat{d}_{\mathcal{X}}} \right). \tag{18}$$

To pass from (15) to (16) we consider only pairs of constellation symbols at minimum distance, and from (16) to (17) we replace each term in the inner sum with its smallest possible value using the fact that the cosine of the angle between $\boldsymbol{t}$ and $\boldsymbol{d}_{i,j}$ is at most $+1$. To pass from (17) to (18), we replace the sum of logarithms by the logarithm of a product, expand it, and then keep the dominant terms (1 and $\mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2 - 2\sqrt{\rho} \|\boldsymbol{t}\| \hat{d}_{\mathcal{X}}}$). Although the error in this last step vanishes at high SNR, we offer no proof for the tightness of the bounds (16) and (17) at high SNR. Nevertheless, the numerical results presented below confirm the usefulness of the bounds (16)–(18).

Theorem 4: For any SNR, the following bounds are valid:

$$\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho) \leq \log(M) + \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} \tilde{g}_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}, \tag{19}$$

$$\mathsf{I}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\rho) \geq \log(M) + \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} \tilde{g}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})\, \mathrm{d}\boldsymbol{t}, \tag{20}$$

where $\tilde{g}_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})$ and $\tilde{g}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})$ are shown in Table I.

Proof: We bound $g_{\boldsymbol{u}}^{\mathrm{CM}}(\boldsymbol{t})$ and $g_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})$ using (16). This used in (6) and (8) gives (19) and (20), respectively. ∎

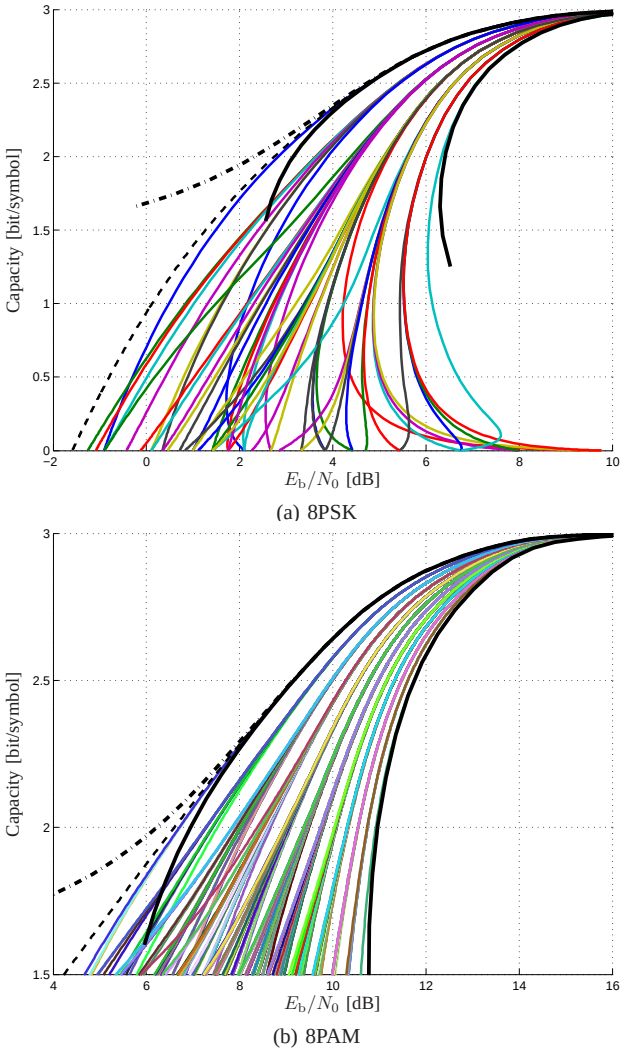
In Fig. 2, we show the results of Theorem 4 and for 8PSK and 8PAM.[4] The lower bound for $\mathsf{I}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\rho)$ is shown for two mappings, namely the ones that maximize/minimize the BICM capacity for high SNR. Both the upper bound for $\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho)$ and the lower bound for $\mathsf{I}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\rho)$ perfectly match the corresponding capacity curves above 2.5 bit/symbol, and thus, the bound in (16) seems to be tight for high SNR.

Theorem 5: For any SNR, the following bounds are valid:

$$\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho) \leq \log(M) - \frac{1}{M\pi^{N/2}} \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} \log \left(1 + A_{\mathcal{X}} \mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2 - 2\sqrt{\rho} \|\boldsymbol{t}\| \hat{d}_{\mathcal{X}}} \right) \mathrm{d}\boldsymbol{t}, \tag{21}$$

$$\mathsf{I}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\rho) \geq (1-m)\log(M) + m\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho) + \frac{1}{M\pi^{N/2}} \int_{\mathbb{R}^N} \mathrm{e}^{-\|\boldsymbol{t}\|^2} \log \left(1 + A_\phi \mathrm{e}^{-\rho \hat{d}_{\mathcal{X}}^2 - 2\sqrt{\rho} \|\boldsymbol{t}\| \hat{d}_{\mathcal{X}}} \right) \mathrm{d}\boldsymbol{t}. \tag{22}$$

[4]Theorems 4 and 5 can be implemented using GH quadratures.



(a) 8PSK

(b) 8PAM

Figure 2. CM capacity (black dashed line) and BICM capacity for all the possible mappings (colored lines) for 8PSK (a) and 8PAM (b). The black dashed-dotted line is the upper bound for $\mathsf{I}_{\boldsymbol{u}}^{\mathrm{CM}}(\rho)$ in (19) and the solid black lines are the lower bound for $\mathsf{I}_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\rho)$ in (20) for two mappings (the ones giving the largest/smallest BICM capacity in the high-SNR regime). Note that the vertical ranges are different.

Proof: Combining the bound in (18) with (6) gives (21). Combining the bound in (18) with (8) and replacing the remaining double summation of logarithms in $g_{\boldsymbol{u},\phi}^{\mathrm{BI}}(\boldsymbol{t})$ by the logarithm of a double product, expanding them, and then keeping the dominant terms gives (22). ∎

Theorem 5 shows that the relevant parameter for the upper bound of the AMI in (21) is the minimum distance of the constellation and its multiplicity. A similar conclusion was drawn in [22, Sec. 3.2.2], in [23, Sec. II-C], and also indirectly in [24, Theorem 4]. On the other hand, Theorem 5 also suggests that the high-SNR behavior of the BICM capacity is captured by the number of minimum distances in the subconstellations $\mathcal{X}_{k,u}$, via the parameter A_ϕ.

The parameter A_ϕ can be regarded as twice the total number of bits being equal between the labelings of pairs of constellation symbols at minimum distance. Therefore, the

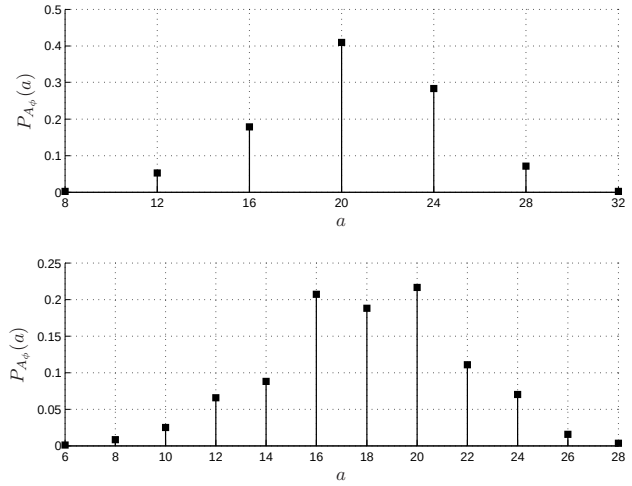


Figure 3. PMF of A_ϕ for 8PSK (top) and 8PAM (bottom) obtained via an exhaustive search. The extreme values are $P_{A_\phi}(32) = 12/7!$ and $P_{A_\phi}(28) = 144/8!$, respectively.

quantity $A_\phi/A_\mathcal{X}$ is the average number of bits being equal between the labelings of the symbols in such pairs. Since $A_\phi/A_\mathcal{X}$ plus the average number of bits being different add up to m, and the average number of different bits is at least one, $A_\phi/A_\mathcal{X} \leq m - 1$.[5] A mapper that meets this bound with equality is a Gray code. For some constellations, there exist a multitude of Gray codes, whereas for others, none at all. We conjecture that for any constellation, Gray codes, if they exist, yield a higher BICM capacity than any non-Gray code.

In Fig. 3 we show the PMF of A_ϕ for 8PSK and 8PAM for all the possible mappings. This figure shows that for 8PSK there exist 7 different values for A_ϕ, which explains the 7 classes of mappings in Fig. 2 (a). For 8PAM there are 12 different values for A_ϕ, which again explains the 12 classes of mappings in Fig. 2 (b). In both cases the highest values are achieved by Gray codes (there is one for 8PSK and three for 8PAM [13, eqs. (15)–(17)]), which supports our conjecture.

V. Conclusions

In this paper we provide ready-to-use formulas for the computation of the CM and BICM capacities based on GH quadratures and used them to study the BICM capacity. The numerical results suggests that the BICM capacity for high SNR can be classified into a limited number of classes with similar behavior, of which the best class consists of all Gray codes. We conjecture that that this is a general property of any constellation for which Gray codes exist. This conjecture is supported by numerical examples and analytical bounds, although a rigorous proof is yet lacking.

[5]The quantity $m - A_\phi/A_\mathcal{X}$ is proportional to $w_0(1)$ in [20] and $\bar{d}(1)$ in [15]. Minimizing either $w_0(1)$ or $\bar{d}(1)$—which translates into a minimization of the average bit-error probability for uncoded transmission—is equivalent to a maximization of A_ϕ.

References

[1] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.

[2] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.

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

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

[5] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Foundations and Trends in Communications and Information Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[6] H. Chen and M. Chen, "Capacity of the distributed antenna systems over shadowed fading channels," in *IEEE Vehicular Technology Conference (VTC-Spring)*, Barcelona, Spain, Apr. 2009.

[7] T. Q. Duong and H.-J. Zepernick, "On the ergodic capacity of cooperative spatial multiplexing systems in composite channels," in *IEEE Radio and Wireless Symposium (RWS)*, San Diego, CA, USA, Jan. 2009.

[8] P. E. McIllree, "Calculation of channel capacity for M-ary digital modulation signal sets," in *IEEE Singapore International Conference on Information Engineering*, Singapore, Sep. 1993.

[9] ——, "Channel capacity calculations for M-ary N-dimensional signal sets," Master's thesis, The University of South Australia, Adelaide, Australia, 1995.

[10] D. Torrieri and M. Valenti, "Rapidly-converging series representations of a mutual-information integral," *ISRN Communications and Networking*, vol. 2011, 2011, article ID 546205.

[11] L. Szczecinski, F.-K. Diop, and M. Benjillali, "On the performance of BICM with mapping diversity in hybrid ARQ," *Wiley Journal Wireless Comm. and Mob. Comput.*, pp. 963–972, Sep. 2008.

[12] C. Nannapaneni, M. C. Valenti, and X. Xiang, "Constellation shaping for communication channels with quantized outputs," in *Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, Mar. 2011.

[13] 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.

[14] 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/.

[15] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "On the optimality of the binary reflected Gray code," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3170–3182, Dec. 2004.

[16] 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.

[17] 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.

[18] E. Agrell and A. Alvarado, "Optimal signal sets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, 2011 (to appear), available at http://arxiv.org/abs/1001.4548.

[19] R. F. Churchhouse, Ed., *Handbook of applicable mathematics. Vol 3: Numerical Methods*. John Wiley & Sons, 1981.

[20] F. Brännström and L. K. Rasmussen, "Classification of unique mappings for 8PSK based on bit-wise distance spectra," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, pp. 1131–1145, Mar. 2009.

[21] F. Brännström, "Convergence analysis and design of multiple concatenated codes," Ph.D. dissertation, Chalmers University of Technology, Göteborg, Sweden, Mar. 2004.

[22] R. De Gaudenzi, A. Guillén i Fàbregas, and A. Martinez, "Turbo-coded APSK modulations design for satellite broadband communications," *Wiley International Journal of Satellite Communications and Networking*, vol. 24, no. 4, pp. 261–281, Jul. Aug. 2006.

[23] F. Pérez-Cruz, M. R. D. Rodriguez, 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.

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