On the capacity of BICM with QAM constellations (Invited Paper)

Alex Alvarado, Erik Agrell, and Arne Svensson*
Department of Signals and Systems
Chalmers University of Technology
Gothenburg, Sweden
{alex.alvarado,agrell,arnes}@chalmers.se

ABSTRACT

In this tutorial paper we analyze the capacity of bit-interleaved coded modulation (BICM) with quadrature amplitude modulation (QAM) constellations, and we pay special attention to different bit-to-symbol labeling strategies. The relation between the BICM capacity and the capacity of other CM schemes such as trellis coded modulation (TCM) and multilevel codes (MLC) is analyzed. Motivated by the fact that for BICM with some particular labelings, the same E_b/N_0 maps to more than one BICM capacity value, we study the relation between the capacity and E_b/N_0. In particular, we present some analytical results on this relation, and we also give an intuitive explanation for the somehow contradictory behavior of these curves.

Categories and Subject Descriptors: E.4 Coding and Information Theory,Error control codes.

General Terms: Theory.

Keywords: BICM, Binary Labeling, Channel Capacity, Coded Modulation, Gray Code, MLC, Quadrature Amplitude Modulation, TCM.

1. INTRODUCTION

The problem of reliable transmission of binary information through a noisy channel dates back to Shannon's work in 1948 [1]. After he introduced the famous capacity formula for the AWGN channel, the problem of designing a system that operates close to that limit has been one of the most important and challenging problems in information/communication theory. Probably, one of the simplest ways of approaching capacity is by using binary signaling (e.g., BPSK) and a binary channel encoder of rate $0 \leq R_c \leq 1$ that corrects errors caused by the channel. In this case, by changing the rate of the encoder, the bit rate can be modified at expense of a higher/lower error correction capability. In such a system the maximum number of bits per channel symbol is bounded by one, and therefore, it is not spectrally efficient for good channel conditions. One straightforward answer to the question of how to efficiently transmit more than 1 bit per symbol is a coded modulation (CM) scheme, where the channel encoder is connected to a modulator where several bits are associated to one channel symbol. What is not straightforward is how to construct such a system, that operates close to the capacity limit, with a reasonably low complexity.

In 1974 Massey proposed the idea of jointly design the channel encoder and modulator [2], which inspired Ungerboeck's trellis coded modulation (TCM) [3], and Imai and Hirakawa's multilevel coding (MLC) [4]. Since both TCM and MLC aim to maximize a Euclidean distance measure, they perform very well over the AWGN channel. However, their performance in fading channels was rather poor.

In 1992 Zehavi introduced the so-called bit-interleaved coded modulation (BICM) [5], which is simply a serial concatenation of a channel encoder, a bit-level interleaver, and a memoryless mapper. BICM aims to increase the code diversity—the key performance measure in fading channels—and therefore, outperforms previous CM schemes in this scenario. When compared to TCM, BICM decreases the minimum Euclidean distance, and consequently, it is suboptimal for the AWGN channel. Nevertheless, since this decrease is only marginal [6], BICM is very robust to variations of the channel characteristics. BICM is very attractive from an implementation point view because of its flexibility, i.e., the channel encoder and the modulator can be selected independently, somehow breaking Massey's joint design paradigm. BICM is nowadays a *de facto* standard, and it is used in most of the existing wireless systems, e.g., HSDPA, IEEE 802.11a/g, IEEE 802.16, DVB-S2, etc.

In this paper we analyze the BICM capacity, its relation with the capacity of TCM and MLC, and we pay special attention to to different bit-to-symbol labeling strategies. We also analyze the BICM capacity vs. E_b/N_0, and we show that for some particular labelings, this relation is not a function, i.e., the same E_b/N_0 maps to more than one capacity value. Since at a first glance the intuition usually fails when trying to understand these figures, we give some analytical and intuitive explanations for their behavior.

The paper is organized as follows. In Sec. 2 we introduce the notation, some definitions, and the system model. In Sec. 3 we analyze the capacity of BICM and its relation with TCM and MLC, and in Sec. 4 we analyze the BICM capacity vs. E_b/N_0. In Sec. 5 the conclusions are drawn.

*Research supported by the Swedish Research Council, Sweden (under research grant #2006-5599).

IWCMC'09, June 21-24, 2009, Leipzig, Germany

2. PRELIMINARIES AND SYSTEM MODEL

Hereafter we use lower case letters x to denote a scalar, and boldface letters $\mathbf{x}$ to denote a vector of scalars. Capital letters X denote random variables, $\mathbb{P}(\cdot)$ denotes probability, $\mathbb{E}[\cdot]$ denotes expectation, and $p_{\mathbf{X}}(\mathbf{x})$ denotes the probability density function (pdf) of the random vector $\mathbf{X}$. Blackboard bold letters $\mathbb{X}$ represent matrices or vectors.

2.1 Binary Labelings

Definition 1 (Binary labeling). *A binary labeling $\mathbb{L}$ of order $m \in \mathbb{Z}^+$ is a sequence of $M = 2^m$ distinct binary codewords, $\mathbb{L} = [\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{M-1}]$, where each $\mathbf{c}_i \in \{0,1\}^m$. A rectangular binary labeling $\mathbb{L}$ of order $(m_1, m_2) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ consists of all codewords in $\{0,1\}^{m_1+m_2}$, arranged in a matrix of dimension $M_1 = 2^{m_1}$ by $M_2 = 2^{m_2}$.*

Definition 2 (Labeling expansion). *To generate a labeling $\mathbb{L}_m$ from a labeling $\mathbb{L}_{m-1} = [\mathbf{c}_0, \dots, \mathbf{c}_{M/2-1}]$, do the following. Repeat each codeword once to obtain a new vector $[\mathbf{c}_0, \mathbf{c}_0, \dots, \mathbf{c}_{M/2-1}, \mathbf{c}_{M/2-1}]$, and then obtain $\mathbb{L}_m$ by adding one coordinate, from the right, taken form the length-M vector $[0\,1\,1\,0\,0\,1\,1\,0 \dots 0\,1\,1\,0]$.*

Definition 3 (Labeling repetition). *To generate a labeling $\mathbb{L}_m$ from a labeling $\mathbb{L}_{m-1} = [\mathbf{c}_0, \dots, \mathbf{c}_{M/2-1}]$, do the following. Repeat the labeling $\mathbb{L}_{m-1}$ once to obtain a new sequence of M vectors $[\mathbf{c}_0, \dots, \mathbf{c}_{M/2-1}, \mathbf{c}_0, \dots, \mathbf{c}_{M/2-1}]$. Add an extra coordinate to each codeword from the left. This extra coordinate is zero (or one) for the first (or last) $M/2$ vectors.*

In this paper we are particularly interested in the binary reflected Gray code (BRGC) [7,8], and in the natural binary code (NBC). We also introduce a new mapping denoted binary semi-Gray code (BSGC) which has some interesting properties that will be analyzed later in the paper.

Definition 4 (Binary reflected Gray code). *The BRGC of order m, denoted by $\mathbb{G}_m$, is generated by $m-1$ recursive expansions of the trivial labeling $\mathbb{L}_1 = [0\,1]$, for any $m \geq 1$.*

Definition 5 (Natural binary code). *The NBC of order m, denoted by $\mathbb{N}_m$, is generated by $m-1$ recursive repetitions of the trivial labeling $\mathbb{L}_1 = [0\,1]$, for any $m \geq 1$.*

Definition 6 (Binary semi-Gray code). *The BSGC of order m, denoted by $\mathbb{S}_m$, for any $m \geq 3$, is generated based on the BRGC of order m as follows. To generate $\mathbb{S}_m$, take $\mathbb{G}_m$ and replace the first coordinate of $\mathbb{S}_m$ by the modulo-2 addition of the first and last coordinates of $\mathbb{G}_m$.*

Definition 7 (Rectangular BRGC). *The rectangular BRGC of order (m_1, m_2), denoted by $\mathbb{G}_{m_1,m_2}$, is generated by the direct product of two BRGC of order m_1 and m_2, i.e., $\mathbb{G}_{m_1,m_2} = \mathbb{G}_{m_1} \times \mathbb{G}_{m_2}$. This definition applies also to rectangular NBC ($\mathbb{N}_{m_1,m_2}$) and rectangular BSGC ($\mathbb{S}_{m_1,m_2}$).*

Example 1 (Binary labelings $\mathbb{G}_3$, $\mathbb{N}_3$, and $\mathbb{S}_3$).

$$\mathbb{G}_3 = \begin{bmatrix}000 & 001 & 011 & 010 & 110 & 111 & 101 & 100\end{bmatrix}$$
$$\mathbb{N}_3 = \begin{bmatrix}000 & 001 & 010 & 011 & 100 & 101 & 110 & 111\end{bmatrix}$$
$$\mathbb{S}_3 = \begin{bmatrix}000 & 101 & 111 & 010 & 110 & 011 & 001 & 100\end{bmatrix}$$

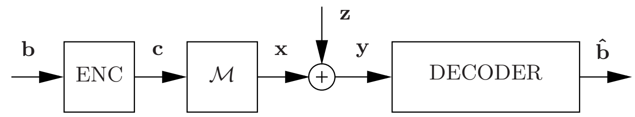

Figure 1: A CM scheme: A channel encoder, a mapper, a discrete-time memoryless AWGN channel, and the decoder.

2.2 System Model

In this paper we analyze coded modulation schemes as the one shown in Fig. 1. At a given time n, k_c information bits are passed to the rate $R_c = k_c/m$ binary channel encoder (ENC), which generates a length-m codeword[1] $\mathbf{c}_n = [c_{n,0}, \dots, c_{n,m-1}]$. This codeword is mapped to a random symbol $\mathbf{X} = \mathbf{x}_n$ in an N-dimensional Euclidean space using a memoryless mapping rule $\mathcal{M} : \{0,1\}^m \to \mathcal{X}$, where $\mathbf{x}_n = [x_{n,1}, \dots, x_{n,N}] \in \mathcal{X}$ and the input alphabet $\mathcal{X}$ is subject to a constraint on the average symbol energy $E_s = \mathbb{E}[||\mathbf{X}||^2]$.

We consider transmissions over the equivalent discrete-time memoryless additive white Gaussian noise (AWGN) channel, with output alphabet $\mathcal{Y} = \mathbb{R}^N$, i.e., $\mathbf{Y} = \mathbf{X} + \mathbf{Z}$, where $\mathbf{Z}$ is a circularly symmetric Gaussian noise with zero mean and variance $N_0/2$ in each dimension. At the receiver's side, and based on the channel observation, the decoder generates an estimate of the information bits $\hat{\mathbf{b}}$.

Each constellation symbol conveys $k_c = R_c m$ information bits, thus, the relation between E_s and the average information bit energy E_b is given by $E_\mathrm{s} = k_c E_\mathrm{b}$. Using this, we can write

$$\frac{E_\mathrm{s}}{N_0} = k_c \frac{E_\mathrm{b}}{N_0}, \tag{1}$$

which will be important for the analysis in Sec. 4.

2.3 AMI and Channel Capacity

In this subsection, we assume that $\mathbf{X}$ can be selected continuously, i.e., $\mathcal{X} = \mathbb{R}^N$, which upperbounds the performance of finite input alphabets. The input symbols are selected with pdf $p_{\mathbf{X}}(\mathbf{x})$ and the conditional channel transition pdf is $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$.

Definition 8 (Average mutual information). *The average mutual information (AMI) between the random variables $\mathbf{X}$ and $\mathbf{Y}$ is defined as*

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}) &\triangleq \mathbb{E}\left[\log_2 \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}|\mathbf{X})}{p_{\mathbf{Y}}(\mathbf{Y})}\right] \\ &= \int_{\mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) \int_{\mathcal{Y}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \log_2 \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})}\, d\mathbf{y}\, d\mathbf{x}. \end{aligned} \tag{2}$$

Definition 9 (Channel capacity). *The channel capacity of a memoryless channel is defined as the maximum AMI between its input and output*

$$\mathcal{C}\left(\frac{E_\mathrm{s}}{N_0}\right) \triangleq \max_{p_{\mathbf{X}}(\mathbf{x})} \{I(\mathbf{X};\mathbf{Y})\}, \tag{3}$$

[1]The encoder has in general a memory. It is implemented as a convolutional code or an (n,k) block code, for any (possibly large) parameters k and n such that $k/n = R_c$. In this case, $k_c = km/n$ does not have to be an integer.

where the maximization is over all possible input distributions.

The capacity in (3) has units of [bit/symbol], and it is an upper bound on the number of bits per symbol that can be reliably transmitted through the channel. Shannon theory states that it is simply not possible to transmit information reliably above this fundamental limit, i.e.,

$$k_c \leq \mathcal{C}\left(\frac{E_\mathrm{s}}{N_0}\right) = \mathcal{C}\left(k_c \frac{E_\mathrm{b}}{N_0}\right). \tag{4}$$

Since the noise is circularly symmetric, the transmission of $\mathbf{X}$ can be considered as a transmission through N parallel Gaussian channels, i.e., $Y_i = X_i + Z_i$, where $Z_i \sim \mathcal{N}(0, N_0/2)$, for $i = 1, \ldots, N$.

Definition 10 (AWGN capacity). *The channel capacity of the AWGN channel is given by [9, Sec. 9.4]*

$$\mathcal{C}^{\mathrm{AW}}\left(\frac{E_\mathrm{s}}{N_0}\right) \triangleq \frac{N}{2}\log_2\left(1 + \frac{E_\mathrm{s}/N}{N_0/2}\right), \tag{5}$$

which is obtained when $X_i \sim \mathcal{N}(0, E_\mathrm{s}/N)$ in each dimension.

3. CAPACITY OF CODED MODULATION SYSTEMS

In this section we analyze three different ways of constructing a CM scheme: TCM, MLC, and BICM. We pay special attention to their capacities, and how they are related.

3.1 TCM and MLC-MSD

For practical reasons, here we restrict our attention to discrete alphabets with cardinality $|\mathcal{X}| = M = 2^m$, where $\mathcal{X} \subset \mathbb{R}^2$ ($N = 2$) formed by the direct product of two equidistant PAM constellations $\mathcal{X}_{\mathrm{PAM}} = \{\pm(\sqrt{M}-1), \pm(\sqrt{M}-3), \ldots, \pm 1\}$, i.e., $\mathcal{X}_{\mathrm{QAM}} = \mathcal{X}_{\mathrm{PAM}} \times \mathcal{X}_{\mathrm{PAM}}$. Moreover, we restrict the input distribution of the symbols to be *uniform*, i.e., $\mathbb{P}(\mathbf{x}) = 1/M$. Under the previous assumptions, we define the so-called uniform capacity as follows.

Definition 11 (Uniform capacity).

$$\mathcal{C}^{\mathrm{UN}}\left(\frac{E_\mathrm{s}}{N_0}\right) \triangleq I(\mathbf{X}; \mathbf{Y}) \tag{6}$$

$$= I(C_0, \ldots, C_{m-1}; \mathbf{Y}), \tag{7}$$

where $\mathbf{C} = (C_0, \ldots, C_{m-1})$ are the binary random variables representing the bits in the codewords in Fig. 1.

To pass from (6) to (7), we used the fact that the mapping rule between $\mathbf{X}$ and $\mathbf{C}$ is one-to-one. Since $\mathcal{X}$ is discrete, to compute the AMI in (2), the outer integral should be replaced by a summation.

Using the chain rule of mutual information [9, Sec. 2.5], the UN capacity in (7) can be rewritten as

$$\mathcal{C}^{\mathrm{UN}}\left(\frac{E_\mathrm{s}}{N_0}\right) = \sum_{k=0}^{m-1} I(C_k; \mathbf{Y} | C_0, \ldots, C_{k-1}) \tag{8}$$

$$= \sum_{k=0}^{m-1} \mathcal{C}_k\left(\frac{E_\mathrm{s}}{N_0}\right), \tag{9}$$

where $\mathcal{C}_k(E_\mathrm{s}/N_0) \triangleq I(C_k; \mathbf{Y}|C_0, \ldots, C_{k-1})$ is a *bit level* capacity which represents the maximum rate that can be used at the $(k+1)$-th bit position, given a perfect knowledge of the other k bits. It is important to note that the UN capacity *does not* depend on the labeling used. Different labelings will produce different values of $\mathcal{C}_k(E_\mathrm{s}/N_0)$ in (9), but the overall sum will remain constant.

The UN capacity can be achieved by a joint design of the encoder and mapper, for example by TCM [3], or by MLC [4, 10] with multistage decoding (MSD). MLC-MSD is in fact a direct application of (8), i.e., m parallel encoders are used, each of them having a rate $R_{c_k} = \mathcal{C}_k(E_\mathrm{s}/N_0)$. At the receiver's side, the first bit level is decoded and the decisions are passed to the second decoder, which then passes the decisions to the third decoder, and so on.

3.2 BICM and MLC-PDL

In a BICM system [5,6] shown in Fig. 2, a bit level interleaver is placed between the encoder and the mapper. At the receiver's side, the demapper computes soft information on the coded bits, which are then deinterleaved and passed to the channel decoder. The demapper computes the a posteriori L-values for the k-th bit in the symbol as

$$l'_k(\mathbf{y}) \triangleq \log\frac{\mathbb{P}(\mathbf{y}|c'_k = 1)}{\mathbb{P}(\mathbf{y}|c'_k = 0)} + \log\frac{\mathbb{P}(c'_k = 1)}{\mathbb{P}(c'_k = 0)}, \tag{10}$$

where the second term in (10) represents the *a priori* information. When a non-iterative BICM receiver is used, the demapper has no a priori information on the coded bits, and consequently, the L-values are calculated as

$$l'_k(\mathbf{y}) = \sum_{u=0}^{1}(-1)^{u+1}\log\sum_{\mathbf{x}\in\mathcal{X}_u^k}\exp\left(-\frac{||\mathbf{y}-\mathbf{x}||^2}{N_0}\right), \tag{11}$$

where $\mathcal{X}_u^k$ are all $\mathbf{x} \in \mathcal{X}$ labeled with $u \in \{0, 1\}$ at position k.

Using the equivalent channel model of [6], m parallel binary-input soft-ouput channels can be defined. If we define $\mathcal{C}'_k(E_\mathrm{s}/N_0) \triangleq I(C_k; \mathbf{Y})$ as the unconditional bit level capacity, and assuming ideal interleaving, each equivalent channel is randomly selected, and therefore, the BICM capacity is equal to the sum of $\mathcal{C}'_k(E_\mathrm{s}/N_0)$ [6].

Definition 12 (BICM capacity).

$$\mathcal{C}_{\mathbb{L}}^{\mathrm{BI}}\left(\frac{E_\mathrm{s}}{N_0}\right) \triangleq \sum_{k=0}^{m-1}\mathcal{C}'_k\left(\frac{E_\mathrm{s}}{N_0}\right). \tag{12}$$

Unlike the UN capacity, the labeling strongly affects the BICM capacity in (12). Note also that the BICM capacity is equivalent to the capacity achieved by MLC with (suboptimal) parallel decoding of the individual bit levels (PDL), i.e., when no information is passed between the m decoders [10]. In BICM the bits are treated as independent, therefore, BICM is somehow analogous to MLC-PDL. The differences are that BICM uses only one encoder, and that in BICM the equivalent channels are not used in parallel, but time multiplexed.

The capacities defined above can be found in the literature under different names. The UN capacity is called joint capacity in [11], (constellation) constrained capacity in [12], or coded modulation capacity in [6]. The BICM capacity is referred as parallel decoding capacity in [11], or receiver constrained capacity in [12]. Furthermore, we recognize that we use the term "capacity" in a broad sense, i.e., without optimizing the input distribution. Also note that the UN capacity and the BICM capacity can be increased if the uniform

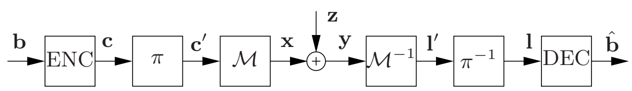

Figure 2: A BICM scheme: A channel encoder, an interleaver, a mapper, an AWGN channel, and the inverse processes at the receiver's side.

input distribution constraint is relaxed, i.e., if the symbols are transmitted with different probabilities (or using equally likely but non-equally spaced constellations), which is usually referred in the literature as *constellation shaping*.

It was demonstrated in [6] that the BICM capacity is less than or equal to the CM capacity, i.e., from a capacity point of view, BICM is suboptimal. This can be intuitively understood based on the definitions of $\mathcal{C}_k(E_\mathrm{s}/N_0)$ and $\mathcal{C}'_k(E_\mathrm{s}/N_0)$, i.e., in BICM the L-values of the coded bits within a received symbol are calculated independently of the other $m-1$ bits, and consequently, some performance loss should be expected.

In Fig. 3 we show the BICM capacity and the UN capacity for 64-QAM and different labelings. We use 64-QAM as a representative case of high order modulations. From these curves we can see that the difference between the UN capacity and the BICM capacity is small when the BRGC is used (above 1 bit/symbol). On the other hand, the gap between the UN capacity and the BICM capacity for other mappings (NBC or BSGC) can be quite large. A very important question here is what is the optimum labeling from a capacity maximization point of view. Once this question is answered, approaching the fundamental limit will depend only on a good design of the channel encoder/decoder. Caire *et al.* conjectured the optimality of the BRGC, however, from this figure it is clear that for low rates (below 1 bit/symbol) the BRGC is not optimum anymore. This problem has been partially analyzed in [13], where an exhaustive search for M-PAM constellations was performed. However, the search was carried out only up to $M=8$, as for $M>8$ the exhaustive search has a prohibitive complexity.

4. THE BICM CAPACITY IS NOT A FUNCTION OF E_b/N_0

In this section we analyze a transformation of the capacity presented in [14], where the BICM capacity was plotted vs. E_b/N_0. The motivation for analyzing this transformation is that when BICM is considered, it is possible that a given E_b/N_0 maps to more than one capacity value, and thus, the capacity curves are not in general a function of E_b/N_0. Similar capacity curves have appeared in different contexts such as analysis of linear precoding for BICM with iterative demapping and decoding [15], or capacity of incoherent MPSK [16] or FSK [17] channels. Moreover, we note that the analysis presented in this section can be also applied to cutoff rate curves.

To avoid confusions, we start by giving a formal definition of a *function*. We adopt this name following its most common definition in the context of differential and integral calculus. This kind of functions are sometimes called single-valued functions, for more details see [18, Sec. 4.2]. A function is a rule of correspondence that associates a real number $y=f(x)$ with each given real number x (the argument), under the restriction that only one value of the function corresponds to the value of the argument. Using this definition, $f_1(x)=x^2$ is a function of $x\in\mathbb{R}$, however, $f_2(x)=\pm\sqrt{x}$ with $x\in\mathbb{R}^+$ is not.

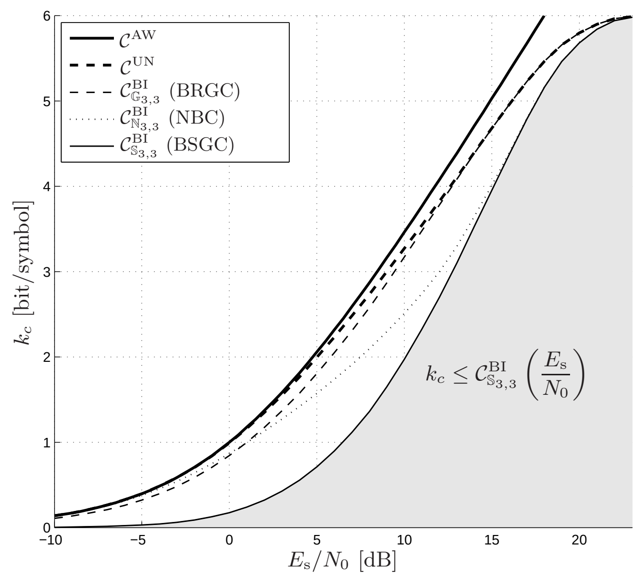

Figure 3: UN capacity and BICM capacity for 64-QAM using different labelings defined in Sec. 2.1. The shadowed region represents the achievable rates for $\mathcal{C}^{\mathrm{BI}}_{\mathbb{S}_{3,3}}$, cf. (4)**.**

Both Verdú [19, Sec. III] and Martinez *et al.* [14, Sec. I] refer to the relation between the capacity and E_b/N_0 as a function, however, we avoid using that term since it is in general not true. We recognize however that [19] did not analyze BICM, and therefore, this effect did not appear, and that most of the labelings analyzed in [14] did not produce this effect either. Moreover, Martinez *et al.* clearly noted this effect since they mentioned that there exist communication schemes for which the minimum value of E_b/N_0 is achieved at nonzero rates.

Since the capacity is a strictly increasing function of E_s/N_0, it has an inverse denoted by $\mathcal{C}^{-1}(k_c)$. From the inequality in (4), we have that

$$\frac{E_\mathrm{b}}{N_0} \ge \frac{\mathcal{C}^{-1}(k_c)}{k_c} \triangleq f(k_c), \tag{13}$$

which gives a lower bound on E_b/N_0 for reliable transmission of information at rate k_c.

Since analytical expressions for the inverse function of the capacity are not available, expressions for $f(k_c)$ in (13) are rare in the literature. One well-known exception is the capacity of the Gaussian channel given by (5), where

$$f^{\mathrm{AW}}(k_c) = \frac{N}{2k_c}(2^{2k_c/N}-1), \tag{14}$$

which results in Shannon's well-known fundamental limit

$$\lim_{k_c\to 0} f^{\mathrm{AW}}(k_c) = \left.\frac{E_\mathrm{b}}{N_0}\right|^{\mathrm{AW}}_{k_c=0} = -1.59\ \mathrm{dB}. \tag{15}$$

In Fig. 4 we present the numerical evaluation of $f(k_c)$ in (13) for the same scenarios presented in Fig. 3. From this

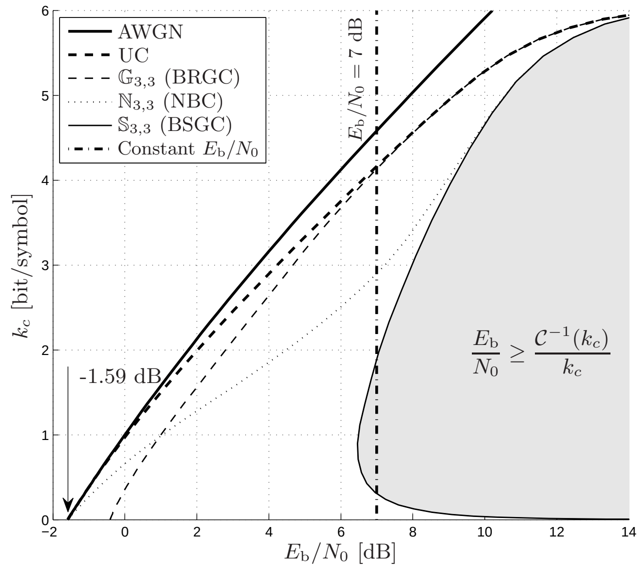

Figure 4: The function $f(k_c)$ for the UC and BICM capacities for 64-QAM. The shadowed region represents the achievable rates given by (13), and the vertical line represents a constant $E_b/N_0 = 7$ dB.

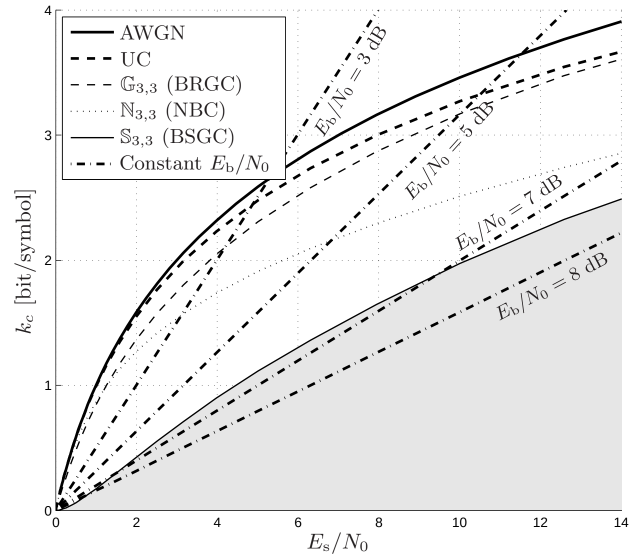

Figure 5: UN capacity and BICM capacity for 64-QAM as a function of E_s/N_0 (linear). The shadowed region is the same as in Fig. 3. The dashed lines are the evaluation of (16) for different values of E_b/N_0.

figure we can see that the UN capacity and the BICM capacity using the NBC achieve $E_b/N_0|^{\text{AW}}_{k_c=0}$. The intersection between the curves for $\mathbb{G}_{3,3}$ and $\mathbb{N}_{3,3}$ can be appreciated at approximately $k_c = 1$ bit/symbol. From this figure is also clear that—asymptotically speaking—Gaussian inputs are not needed to achieve $E_b/N_0|^{\text{AW}}_{k_c=0}$ [19]. In this figure we also present the region given by (13) where reliable communication is possible for 64-QAM using the BSGC. Based on these results, it is very clear that the capacity curves are not, in general, a function of E_b/N_0.

If we analyze the curve for $\mathbb{S}_{3,3}$ in Fig. 4, at a first glance it looks somehow contradictory that for a given E_b/N_0 (e.g., $E_b/N_0 = 7$ dB), there exists an achievable region (between 0.3 and 1.9 bit/symbol), however, there are two regions that are not achievable (lower than 0.3 bit/symbol and higher 1.9 bit/symbol). The intuition usually fails at this point since it is not clear how, by decreasing the information rate, a non-achievable region is reached. One simple argument is that, for a given code operating at an achievable rate for $E_b/N_0 = 7$ dB, k_c can be decreased by simply adding dummy bits that will be discarded at the receiver. In that case, and since the error correcting capabilities of the code have not been modified, it is very counterintuitive to understand how this new scheme could result in a non-achievable region. To understand this, we first present an intuitive explanation, and then we analyze the problem from an analytical point of view.

Starting with the fundamental relation in (1), we note that for a fixed E_b/N_0 (vertical lines in Fig. 4), the relation between k_c and E_s/N_0 is linear

$$k_c = \frac{1}{E_b/N_0}\frac{E_s}{N_0}. \tag{16}$$

In Fig. 5 we present the relation in (16) for different values of E_b/N_0 (note the linear scale for E_s/N_0) together with the capacity curves of Fig. 3. From this figure we can see that (16) results in lines with different slopes, all of them starting at the origin. Independently of the considered scheme, any capacity is zero for $E_s/N_0 = 0$, and therefore, all the capacity curves and the curves from (16) intersect at the origin. From this figure it is also clear how the capacity curve for $\mathbb{S}_{3,3}$—which is not a concave function—intersects the constant E_b/N_0 curves twice for non-zero rates and for sufficiently high E_b/N_0 (above ≈ 6 dB). On the other hand, all the other curves in Fig. 5 are concave functions, and therefore, they intersect the constant E_b/N_0 curves only once for non-zero rates. A concave capacity curve translates into a function $f(k_c)$ in Fig. 4 that has a minimum at $k_c = 0$. However, the converse is not necessarily true; a nonconcave capacity curve may still have a unique minimum at $k_c = 0$.

Now we go back to the counterintuitive example given before. Assume that we operate in the achievable region for a constant $E_b/N_0 = 7$ dB using a rate $R_c = 1/6$ and BICM with $\mathbb{S}_{3,3}$ (consequently $k_c = 1$ bit/symbol). A decrease (or increase) in the rate should be understood as moving to the left (or to the right) *along* the line of $E_b/N_0 = 7$ dB in Fig. 5. From this curve, we can clearly see that, when moving to the left, there will be a point where reliable communication is indeed not possible ($E_s/N_0 \lesssim 1.5$). Moreover, the argument of transmitting dummy bits can be proved wrong as follows. If E_b/N_0 is constant, and the code rate R_c decreases because the use of dummy bits, k_c also decreases. In this situation, (1) dictates that E_s/N_0 will decrease (energy is wasted in transmitting the dummy bits), and therefore, it is not surprising that an unachievable region can be reached.

In the following we present analytical results that explain the results previously discussed.

Theorem 1 (Minimum E_b/N_0 for non-zero rates). *The minimum E_b/N_0 for a non-zero rate $0 < k_c < m$ is given by $f(\check{k}_c)$, where $\check{k}_c$ is one of the solutions of $g(k_c) = 0$,*

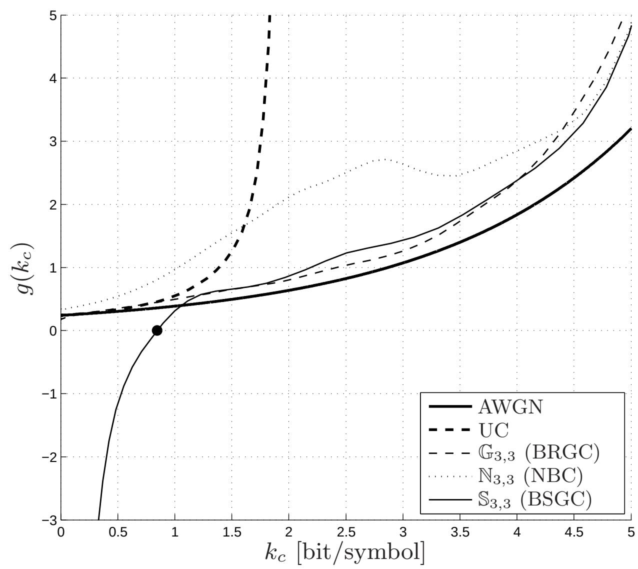

Figure 6: The function $g(k_c)$ in (17) for BICM and different labelings. The filled circle represents the non-zero rate solution of $g(k_c) = 0$.

and

$$g(k_c) \triangleq \frac{1}{k_c}\frac{d\mathcal{C}^{-1}(k_c)}{dk_c} - \frac{\mathcal{C}^{-1}(k_c)}{k_c^2}. \qquad (17)$$

Proof. By solving the derivative of (13) equal to zero. □

Since $f(k_c)$ is in general not known analytically, the function $g(k_c)$ must be numerically evaluated using $\mathcal{C}(E_s/N_0)$. An exception to this is the capacity of the AWGN channel, where $g(k_c)$ can be calculated analytically. Moreover it can be proved that in this case that a minimum E_b/N_0 for non-zero rates does not exist.

Corollary 1. *The minimum E_b/N_0 for the AWGN channel is unique, and it is obtained only for zero-rate transmissions.*

Proof. The derivative of $f^{AW}(k_c)$ in (14) is given by

$$g^{AW}(k_c) = \frac{N + (2k_c \log 2 - N)2^{2k_c/N}}{2k_c^2} = \frac{g_n(k_c)}{g_d(k_c)}, \qquad (18)$$

which results in $\lim_{k_c \to 0} g^{AW}(k_c) = (\log 2)^2/N$. To prove that a minimum for a non-zero rate does not exit, and since $g_d(k_c) > 0$ for $k_c > 0$, we only need to prove that $g_n(k_c) > 0$. Since $\lim_{k_c \to 0} g_n(k_c) = 0$, we simply need to prove that the first derivative of $g_n(k_c)$ is strictly positive. This is trivial since $\frac{dg_n(k_c)}{dk_c} = \frac{4}{N} k_c (\log 2)^2 2^{2k_c/N} > 0$. □

In Fig. 6 we present the function $g(k_c)$ in (17). If $g(k_c) = 0$ has at least one solution, the capacity curve will have a minimum for a non-zero rate (shown with a filled circle in Fig. 6 for $\mathbb{S}_{3,3}$). Note also that the BSGC introduced in Sec. 2.1 has an interesting property. Namely, $\lim_{k_c \to 0} g^{\mathbb{S}_{3,3}}(k_c) = -\infty$, and consequently, $\lim_{k_c \to 0} f^{\mathbb{S}_{3,3}}(k_c) = +\infty$ (cf. Fig. 6 and Fig. 4 for $\mathbb{S}_{3,3}$ with $k_c = 0$). In this sense, the BSGC is an extremely bad labeling for asymptotically low rates.

It is interesting to note that, in general, $g(k_c) = 0$ could have multiple non-zero rate solutions, and therefore, the curve of capacity vs. E_b/N_0 could have an even more strange behavior than $\mathbb{S}_{3,3}$ in Fig. 4. In other words, the same E_b/N_0 could map to three or more capacity values, and therefore, the function $f(k_c)$ will have multiple local minima/maxima.

5. CONCLUSIONS

In this paper we presented an overview of BICM and we analyzed its relationship with other CM schemes such as TCM and MLC. Curves of the BICM capacity vs. the E_b/N_0 were presented and analyzed for QAM schemes and some selected binary labelings. We have showed that when plotting the BICM capacity curves vs. E_b/N_0, counterintuitive results can be obtained, and therefore, we believe that it is more convenient to plot the capacity curves as a function of E_s/N_0.

It is worth mentioning that one of the most interesting open research problems in this field is to determine the optimum capacity maximizing labeling for BICM, for any constellation size, and any E_s/N_0.

6. 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] J. Massey, "Coding and modulation in digital communications," in *International Zürich Seminar on Digital Communications*, Zürich, Switzerland, Mar. 1974.
[3] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 55–67, Jan. 1982.
[4] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Trans. Inf. Theory*, vol. 23, no. 3, pp. 371–377, May 1977.
[5] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 3, pp. 873–884, May 1992.
[6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, May 1998.
[7] F. Gray, "Pulse code communications," U. S. Patent 2 632 058, Mar., 1953.
[8] E. Agrell, J. Lassing, E. G. Ström, and T. Ottosson, "Gray coding for multilevel constellations in Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 224–235, Jan. 2007.
[9] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, USA: John Wiley & Sons, 2006.
[10] U. Wachsmann, R. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, July 1999.
[11] M. Barsoum, C. Jones, and M. Fitz, "Constellation design via capacity maximization," in *IEEE International Symposium on Information Theory, ISIT 2007*, Nice, France, June 2007, pp. 1821–1825.
[12] F. Schreckenbach, "Iterative decoding of bit-interleaved coded modulation," Ph.D. dissertation, Munich University of Technology, Munich, Germany, 2007.

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

[14] 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.

[15] F. Simoens, H. Wymeersch, and M. Moeneclaey, "Linear precoders for bit-interleaved coded modulation on AWGN channels: Analysis and design criteria," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 87–99, Jan. 2008.

[16] M. Peleg and S. Shamai, "On the capacity of the blockwise incoherent MPSK channel," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 603–609, May 1998.

[17] W. E. Stark, "Capacity and cutoff rate of noncoherent FSK with nonselective Rician fading," *IEEE Trans. Commun.*, vol. COM-33, no. 11, pp. 603–609, Nov. 1985.

[18] G. Korn and T. Korn, *Mathematical Handbook for Scientists and Engineers*. New York, USA: McGraw-Hill, 1961.

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