Bit-Interleaved Coded Modulation in the Wideband Regime

Alfonso Martinez, *Member, IEEE*, Albert Guillén i Fàbregas, *Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, and Frans M. J. Willems, *Fellow, IEEE*

Abstract—**The wideband regime of bit-interleaved coded modulation (BICM) in Gaussian channels is studied. The Taylor expansion of the coded modulation capacity for generic signal constellations at low signal-to-noise ratio (SNR) is derived and used to determine the corresponding expansion for the BICM capacity. Simple formulas for the minimum energy per bit and the wideband slope are given. BICM is found to be suboptimal in the sense that its minimum energy per bit can be larger than the corresponding value for coded modulation schemes. The minimum energy per bit using standard Gray mapping on M-PAM or M^2-QAM is given by a simple formula and shown to approach $-$ 0.34 dB as M increases. Using the low SNR expansion, a general tradeoff between power and bandwidth in the wideband regime is used to show how a power loss can be traded off against a bandwidth gain.**

Index Terms—**Additive white Gaussian noise (AWGN) channel, bit-interleaved coded modulation, coded modulation, Rayleigh fading channel, wideband regime.**

I. Introduction and Motivation

BIT-interleaved coded modulation (BICM) was originally proposed by Zehavi [1] and further elaborated by Caire *et al.* [2] as a practical way of constructing efficient coded modulation schemes over nonbinary signal constellations. Reference [2] defined and computed the channel capacity of BICM under a suboptimal *noniterative* decoder, and compared it to the coded modulation capacity, assuming equiprobable signaling over the constellation. When natural reflected Gray mapping was used, the BICM capacity was found to be near optimal at high signal-to-noise ratio (SNR). Nevertheless, plots of the BICM capacity as a function of the energy per bit for reliable communication (see Fig. 1) reveal the suboptimality of BICM with the noniterative decoder of [1], [2] for low rates, that is in the power-limited or *wideband* regime.

Recent work by Verdú [3] presents a detailed treatment of the wideband regime. He studied the minimum bit energy-to-noise ratio $\frac{E_{\mathrm{b}}}{N_{0\mathrm{min}}}$ for reliable communication and the wideband slope, i. e., the first-order expansion of the capacity for low $\frac{E_{\mathrm{b}}}{N_{0\mathrm{min}}}$, under a variety of channel models and channel state information (CSI) assumptions. These results are obtained by using a second-order expansion of the channel capacity at zero SNR. Furthermore, using these results, he obtained a general tradeoff between data rate, power, and bandwidth in the wideband regime. In particular, Verdú [3] studied the bandwidth penalty incurred by using suboptimal signal constellations in the low-power regime. An implicit assumption of this tradeoff was that the power cannot change together with the bandwidth.

Motivated by the results of Fig. 1 and by Verdú's analysis [3], in this paper we give an analytical characterization of the behavior of BICM in the low-power regime. Studying the behavior of BICM at low rates may prove useful in the design of multirate communication systems, where rate adaptation is carried out by modifying the binary code, while keeping the modulation unchanged. In the process, we derive a number of results of independent interest for coded modulation over the Gaussian channel. In particular, the first two coefficients of the Taylor expansion of the coded modulation capacity for arbitrary signal constellations at zero SNR are derived, and used to obtain the corresponding coefficients for BICM. We also obtain a closed-form expression for the minimum $\frac{E_{\mathrm{b}}}{N_0}$ for BICM using quadrature amplitude modulation (QAM) constellations with natural reflected Gray mapping, and we show that for large constellations it approaches -0.34 dB, resulting in a 1.25-dB power loss with respect to coded modulation. Using these results, we derive the tradeoff between power and bandwidth in the wideband regime that generalizes the results of [3] to capture the effects of changing both power and bandwidth.

This paper is organized as follows. Section II introduces the system model, basic assumptions, and notation. Section III defines the wideband regime, and presents the low-SNR expansion for both coded modulation and BICM. Section IV introduces the general tradeoff between power and bandwidth. Concluding remarks appear in Section V. Proofs of various results are in the Appendices.

II. Model and Assumptions

We consider a complex-valued, discrete-time additive Gaussian noise channel with fading. The kth channel output y_k is given by

$$y_k = h_k \sqrt{\mathsf{SNR}}\, x_k + z_k \tag{1}$$
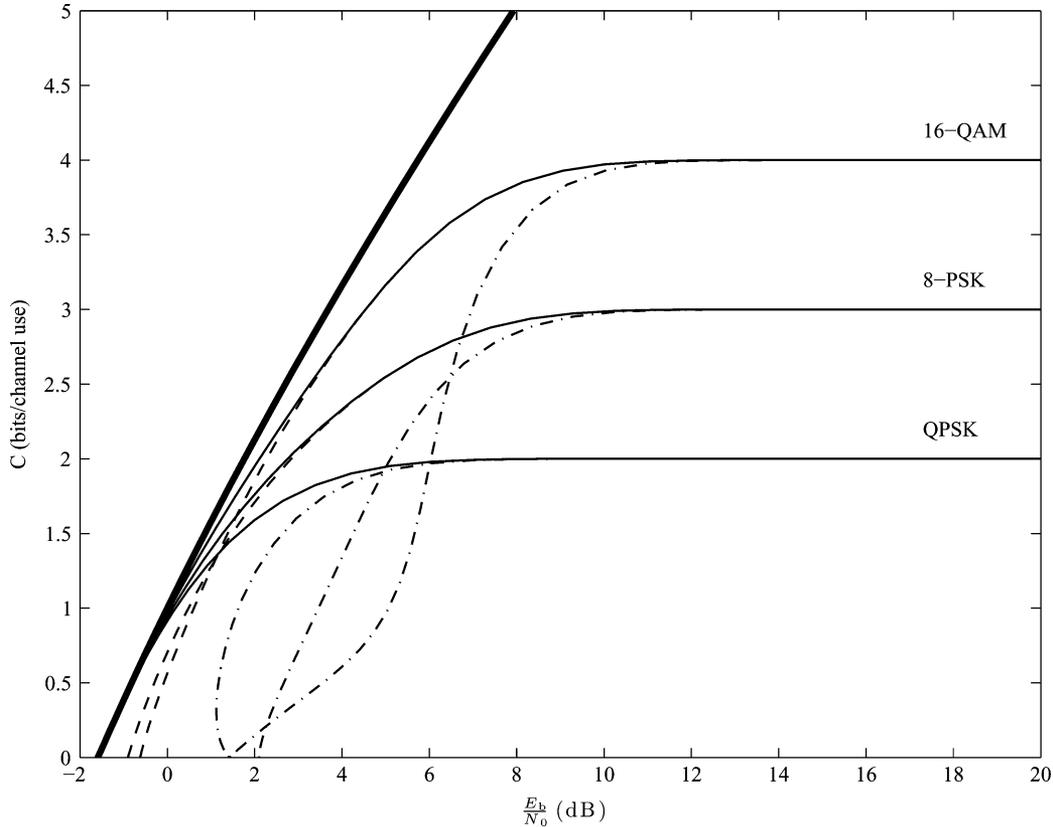
Fig. 1.   Channel capacity (in bits per channel use) as a function of $\frac{E_b}{N_0}$ with memoryless binary labeling and BICM-ML decoding for multiple signal constellations with uniform inputs in the additive white Gaussian noise (AWGN) channel. Gray and set partitioning labeling rules correspond to thin dotted and dashed-dotted lines, respectively. In thick solid lines, the capacity with Gaussian inputs; in thin solid lines the CM channel capacity with uniform inputs (3).

where x_k is the kth channel input, h_k a fading coefficient, and z_k an independent sample of circularly symmetric complex-valued Gaussian noise of unit variance; SNR denotes the average SNR at the receiver. The transmitted, received, noise, and fading samples are realizations of the random variables X, Y, Z, and H. The fading coefficients h_k are independently drawn from a density $p_H(h_k)$ and are assumed known at the receiver. For future use we define the squared magnitudes of the fading coefficients by $\chi_k = |h_k|^2$. For a given fading realization h_k, the conditional output probability density is given by

$$p_{Y|X,H}(y_k|x_k,h_k) = \frac{1}{\pi}e^{-|y_k - h_k\sqrt{\mathsf{SNR}}\,x_k|^2}. \qquad (2)$$

The channel inputs are modulation symbols drawn from a constellation set $\mathcal{X}$ with probabilities $P_X(x)$. We denote the cardinality of the constellation set by $M = |\mathcal{X}|$ and by $m = \log_2 M$ the number of bits required to index a modulation symbol. We define the constrained capacity $\mathrm{C}_{\mathcal{X}}$ (or coded modulation capacity) as the corresponding mutual information between channel input and output, namely

$$\mathrm{C}_{\mathcal{X}}(\mathsf{SNR})$$
$$= -\mathrm{E}\left[\log\left(\sum_{x'\in\mathcal{X}} P_X(x')e^{-|H\sqrt{\mathsf{SNR}}(X-x')+Z|^2+|Z|^2}\right)\right]$$
$$\qquad (3)$$

where the expectation is performed over X, Z, and H. If the symbols are used with equal probabilities, i.e., $P_X(x) = M^{-1}$,

we refer to the constrained capacity as uniform capacity, and denote it by $\mathrm{C}_{\mathcal{X}}^{\mathrm{u}}$.

As we will see later, it proves convenient to consider general constellation sets with arbitrary first and second moments, respectively denoted by $\mu_1(\mathcal{X})$ and $\mu_2(\mathcal{X})$, and given by

$$\mu_1(\mathcal{X}) \triangleq \mathrm{E}[X] = \sum_{x\in\mathcal{X}} x P_X(x)$$
$$\mu_2(\mathcal{X}) \triangleq \mathrm{E}[|X|^2] = \sum_{x\in\mathcal{X}} |x|^2 P_X(x).$$

Practical constellations have zero mean, i.e., $\mu_1(\mathcal{X}) = 0$, and unit energy, that is, $\mu_2(\mathcal{X}) = 1$.

In order to transmit at rates close to the coded modulation capacity, multilevel coding or nonbinary codes are needed [4], [5]. Alternatively, in BICM binary codes are mapped with a binary mapping rule γ onto nonbinary modulations [1], [2]. Caire *et al.* found that BICM with natural reflected Gray mapping and low-complexity noniterative demodulation attains very good performance, close to that of coded modulation with equiprobable signaling [2]. For infinite interleaving, the channel is separated into a set of m parallel independent subchannels, and one defines the so-called BICM capacity, denoted by $\mathrm{C}_{\mathcal{X},\gamma}$, given by

$$\mathrm{C}_{\mathcal{X},\gamma}(\mathsf{SNR}) = \sum_{i=1}^{m} I(B_i;Y) \qquad (4)$$

$$= \sum_{i=1}^{m} \mathrm{E}\left[\log \frac{\sum_{x'\in\mathcal{X}_b^i} e^{-|H\sqrt{\mathsf{SNR}}(X-x')+Z|^2}}{\frac{1}{2}\sum_{x'\in\mathcal{X}} e^{-|H\sqrt{\mathsf{SNR}}(X-x')+Z|^2}}\right] \qquad (5)$$

where B_i denotes the binary input random variable corresponding to the ith parallel channel (see [2] for details), $\mathcal{X}_b^i$ are the sets of constellation symbols with bit b in the ith position of the binary label, and the expectation is performed over all input symbols x in $\mathcal{X}_b^i$ for $b = 0, 1$, and over all possible noise and fading realizations, respectively, Z and H. An equivalent, yet alternative, definition is given by the following.

Proposition 1: The BICM capacity can be expressed as

$$C_{\mathcal{X},\gamma} = \sum_{i=1}^{m} \frac{1}{2} \sum_{b=0,1} \left(C_{\mathcal{X}}^{u} - C_{\mathcal{X}_b^i}^{u} \right) \tag{6}$$

where $C_{\mathcal{X}}^{u}$ and $C_{\mathcal{X}_b^i}^{u}$ are, respectively, the constrained capacities for equiprobable signaling in $\mathcal{X}$ and $\mathcal{X}_b^i$. $\square$

Proof: The proof is given in Appendix A.[1] $\square$

In general, the sets $\mathcal{X}_b^i$ have nonzero mean and nonunit average energy. This result reduces the analysis of the BICM capacity to that of coded modulation over constellation sets with arbitrary first and second moments.

III. WIDEBAND REGIME

In the wideband regime, as defined by Verdú in [3], the energy of a single bit is spread over many channel degrees of freedom, resulting in a low signal-to-noise ratio SNR. It is then convenient to study the asymptotic behavior of the channel capacity as $\mathsf{SNR} \to 0$. In general, the capacity[2] (in nats per channel use) admits an expansion in terms of SNR

$$C(\mathsf{SNR}) = c_1 \mathsf{SNR} + c_2 \mathsf{SNR}^2 + o\left(\mathsf{SNR}^2\right) \tag{7}$$

where c_1 and c_2 depend on the modulation format, the receiver design, and the fading distribution.

Among the several uses for the coefficients c_1 and c_2, Verdú [3] studied the transformation of expansion (7) into a function of the bit-energy-to-noise ratio $\frac{E_b}{N_0}$

$$\frac{E_b}{N_0} = \frac{\mathsf{SNR}}{C \log_2 e}. \tag{8}$$

In linear scale for $\frac{E_b}{N_0}$, one obtains

$$C\left(\frac{E_b}{N_0}\right) = \zeta_0 \left(\frac{E_b}{N_0} - \frac{E_b}{N_{0\lim}}\right) + O\left(\left(\Delta \frac{E_b}{N_0}\right)^2\right) \tag{9}$$

where $\Delta \frac{E_b}{N_0} \triangleq \frac{E_b}{N_0} - \frac{E_b}{N_{0\lim}}$ and

$$\zeta_0 \triangleq -\frac{c_1^3}{c_2 \log^2 2}, \quad \frac{E_b}{N_{0\lim}} \triangleq \frac{\log 2}{c_1}. \tag{10}$$

The parameter ζ_0 is Verdú's wideband slope in linear scale [3]. We avoid using the word minimum for $\frac{E_b}{N_{0\lim}}$, since there exist communication schemes with a negative slope ζ_0, for which the absolute minimum value of $\frac{E_b}{N_0}$ is achieved at nonzero rates. In these cases, the expansion at low power is still given by (9). The derivation of (9) can be found in Appendix B.

A second important use of the coefficients c_1 and c_2 was the analysis of the bandwidth penalty incurred by using suboptimal

constellations in the low-power regime [3]. An implicit assumption in [3] was that the power cannot change together with the bandwidth. In Section IV, we relax this assumption and give a formula for the tradeoff between power penalty and bandwidth penalty and apply it to compare BICM with standard coded modulation.

In the following, we determine the coefficients c_1 and c_2 in the expansion (7) for generic constellations, and use them to derive the corresponding results for BICM. Before proceeding along this line, we note that Theorem 12 of [3] covers the effect of fading. The coefficients c_1 and c_2 for a general fading distribution are

$$c_1 = E[\chi]c_1^{AWGN}, \qquad c_2 = E[\chi^2]c_2^{AWGN} \tag{11}$$

where the coefficients c_1^{AWGN} and c_2^{AWGN} are in absence of fading. Hence, even though we focus only on the AWGN channel, all results are valid for general fading distributions.

A. Coded Modulation

For the unconstrained case, where the capacity is $\log(1 + \mathsf{SNR})$, then $c_1 = 1$ and $c_2 = -\frac{1}{2}$. In [7], Prelov and Verdú determined the coefficients c_1 and c_2 for the so-called proper-complex constellations introduced by Neeser and Massey [8], which satisfy

$$\mu_2'(\mathcal{X}) \triangleq E[X^2] = \sum_{x \in \mathcal{X}} x^2 P_X(x) = 0$$

where $\mu_2'(\mathcal{X})$ is a second-order pseudo-moment, borrowing notation from the paper [8]. The coefficients for coded modulation formats with arbitrary first and second moments are given by the following result.

Theorem 1: Consider coded modulation schemes in the Gaussian channel, over a signal set $\mathcal{X}$ used with probabilities $P_X(x)$. The first two coefficients of the Taylor expansion of the constrained capacity $C_{\mathcal{X}}(\mathsf{SNR})$ around $\mathsf{SNR} = 0$ are given by

$$c_1 = \mu_2(\mathcal{X}) - |\mu_1(\mathcal{X})|^2 \tag{12}$$

$$c_2 = -\frac{1}{2}\left((\mu_2(\mathcal{X}) - |\mu_1(\mathcal{X})|^2)^2 + |\mu_2'(\mathcal{X}) - \mu_1^2(\mathcal{X})|^2\right). \tag{13}$$

When $\mu_1(\mathcal{X}) = 0$ (zero mean) and $\mu_2(\mathcal{X}) = 1$ (unit energy)

$$c_1 = 1, c_2 = -\frac{1}{2}\left(1 + |\mu_2'(\mathcal{X})|^2\right), \tag{14}$$

and the bit-energy-to-noise ratio at zero SNR is $\frac{E_b}{N_{0\lim}} = \log 2$.

Proof: See Appendix C. $\square$

The formula for c_1 is known, and can be found as Theorem 4 of [3]. Also, for proper-complex constellations $c_2 = -\frac{1}{2}$, as found in [7]. The second-order coefficient is bounded by $-1 \leq c_2 \leq -\frac{1}{2}$, the maximum $(c_2 = -1/2)$ being attained when the constellation has uncorrelated real and imaginary parts and the energy is equally distributed among the real and imaginary parts.

Applied to some practical signal constellations with equiprobable symbols, Theorem 1 gives the following corollaries, whose respective proofs are straightforward.

Corollary 1: For uniform M-PSK, $c_2 = -1$ if $M = 2$ and $c_2 = -\frac{1}{2}$ if $M > 2$.

---

[1]This expression has been independently derived in [6].

[2]This capacity may be the coded modulation capacity, or the BICM capacity.

TABLE I
$\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$ AND WIDEBAND SLOPE COEFFICIENTS c_1, c_2 FOR BICM IN AWGN

| | Modulation and Mapping | | | | | |
| | QPSK | | 8-PSK | | 16-QAM | |
	GR	A-GR	GR	SP	GR	SP
c_1	1.000	0.500	0.854	0.427	0.800	0.500
$\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$	0.693	1.386	0.812	1.624	0.866	1.386
$\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$ (dB)	-1.592	1.419	-0.904	2.106	-0.627	1.419
c_2	-0.500	0.250	-0.239	0.005	-0.160	-0.310
ζ_0	4.163	-1.041	5.410	-29.966	6.660	0.839

This result extends Theorem 11.1 of [3], where the result held for quaternary phase-shift keying (QPSK), a simple example of proper-complex constellation.

Corollary 2: When $\mathcal{X}$ represents a mixture of N uniform M_n-PSK constellations for $n = 1, \ldots, N$, $c_2 = -\frac{1}{2}$ if and only if $M_n > 2$ for all rings/subconstellations $n = 1, \ldots, N$.

This applies to amplitude-and-phase-shift keying (APSK) modulations, for instance. In [3] Theorem 11.2 stated the result for mixtures of QPSK constellations.

B. Bit-Interleaved Coded Modulation

First, for fixed label index i and bit value b, let us respectively define the quantities $\mu_1(\mathcal{X}_b^i)$, $\mu_2(\mathcal{X}_b^i)$, and $\mu_2'(\mathcal{X}_b^i)$ as the mean, the second moment, and the average of the squared symbols in the set $\mathcal{X}_b^i$. Then, we have the following.

Theorem 2: Assume a constellation set $\mathcal{X}$ with zero mean and unit average energy. The coefficients c_1 and c_2 for the BICM capacity $C_{\mathcal{X},\gamma}$ are given by

$$c_1 = \sum_{i=1}^{m} \frac{1}{2} \sum_{b} \left| \mu_1(\mathcal{X}_b^i) \right|^2 \tag{15}$$

$$c_2 = \sum_{i=1}^{m} \frac{1}{4} \sum_{b=0,1} \left(\left(\mu_2(\mathcal{X}_b^i) - \left| \mu_1(\mathcal{X}_b^i) \right|^2 \right)^2 \right.$$
$$\left. - \left(1 + |\mu_2'(\mathcal{X})|^2 \right) + \left| \mu_2'(\mathcal{X}_b^i) - \mu_1^2(\mathcal{X}_b^i) \right|^2 \right). \tag{16}$$

Proof: See Appendix D. □

Table I reports the numerical values for the coefficients c_1 and c_2, as well as the bit SNR $\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$ and wideband slope ζ_0 for various cases, namely QPSK, 8-PSK, and 16-QAM modulations and Gray and set partitioning (anti-Gray for QPSK) mappings.

In Fig. 2, the approximation in (9) is compared with the capacity curves. As expected, a good match for low rates is observed. We use labels to identify the specific cases: labels 1 and 2 are QPSK, 3 and 4 are 8-PSK, and 5 and 6 are 16-QAM. Also depicted is the linear approximation to the capacity around $\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$, given by (9). Two cases with Nakagami fading are also included in Fig. 2, which also show good match with the estimate, taking into account that $\mathrm{E}[\chi] = 1$ and $\mathrm{E}[\chi^2] = 1 + 1/\nu$ for Nakagami-ν fading. An exception is 8-PSK with set-partitioning, for which

the approximation is valid for a very small range of rates, since c_2 is positive and very small, which implies a very large slope.

In general, it seems difficult to draw general conclusions for arbitrary mappings from Theorem 2. A notable exception, however, is the analysis under natural reflected Gray mapping.

Theorem 3: For M-PAM and M^2-QAM and natural, binary-reflected Gray mapping, the coefficient c_1 in the Taylor expansion of the BICM capacity $C_{\mathcal{X},\gamma}$ at low SNR is

$$c_1 = \frac{3 \cdot M^2}{4(M^2 - 1)}, \tag{17}$$

and the minimum $\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$ is

$$\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}} = \frac{4(M^2 - 1)}{3 \cdot M^2} \log 2. \tag{18}$$

As $M \to \infty$, $\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}}$ approaches $\frac{4}{3} \log 2 \simeq -0.3424$ dB from below.

Proof: The proof can be found in Appendix E. □

The results for BPSK, QPSK (2-PAM×2-PAM), and 16-QAM (4-PAM×4-PAM), as presented in Table I, match with the theorem.

Somewhat surprisingly, the loss with respect to coded modulation at low SNR is bounded. The loss represents about 1.25 dB with respect to the classical coded modulation (CM) limit, namely, $\frac{E_{\mathrm{b}}}{N_{0\,\mathrm{lim}}} = -1.59$ dB. In the next section, we examine in detail the precise extent to which this loss translates into an equivalent loss in power. We will do so by allowing for simultaneous variations in power and bandwidth and conclude that using BICM over a fixed modulation for a large range of SNR values, where the transmission rate is adjusted by changing the code rate, needs not result in a large loss with respect to more optimal schemes, where both the rate and modulation change. Additionally, this loss can be traded off against a large bandwidth reduction.

IV. BANDWIDTH AND POWER TRADEOFF

In the previous section we computed the first coefficients of the Taylor expansion of the CM and BICM capacities around SNR = 0. In this section, we use these coefficients to determine the tradeoff between power and bandwidth in the low-power regime. We will see how part of the power loss incurred by BICM can be traded off against a large bandwidth reduction.

The data rate transmitted across a Gaussian channel is determined by two physical variables: the power P, or energy per unit time, and the bandwidth W, or number of channel uses per unit time. In this case, the signal-to-noise ratio SNR is given by $\mathsf{SNR} = P/(N_0 W)$, where N_0 is the noise spectral density. Then, the capacity measured in bits per unit time is the natural figure of merit for a communications system. With only a constraint on SNR, this capacity is given by $W \log(1 + \mathsf{SNR})$. For low SNR, we have that

$$W \log \left(1 + \frac{P}{N_0 W} \right) = \frac{P}{N_0} - \frac{P^2}{2 N_0^2 W} + \mathrm{O}\left(\frac{P^3}{N_0^3 W^2} \right). \tag{19}$$
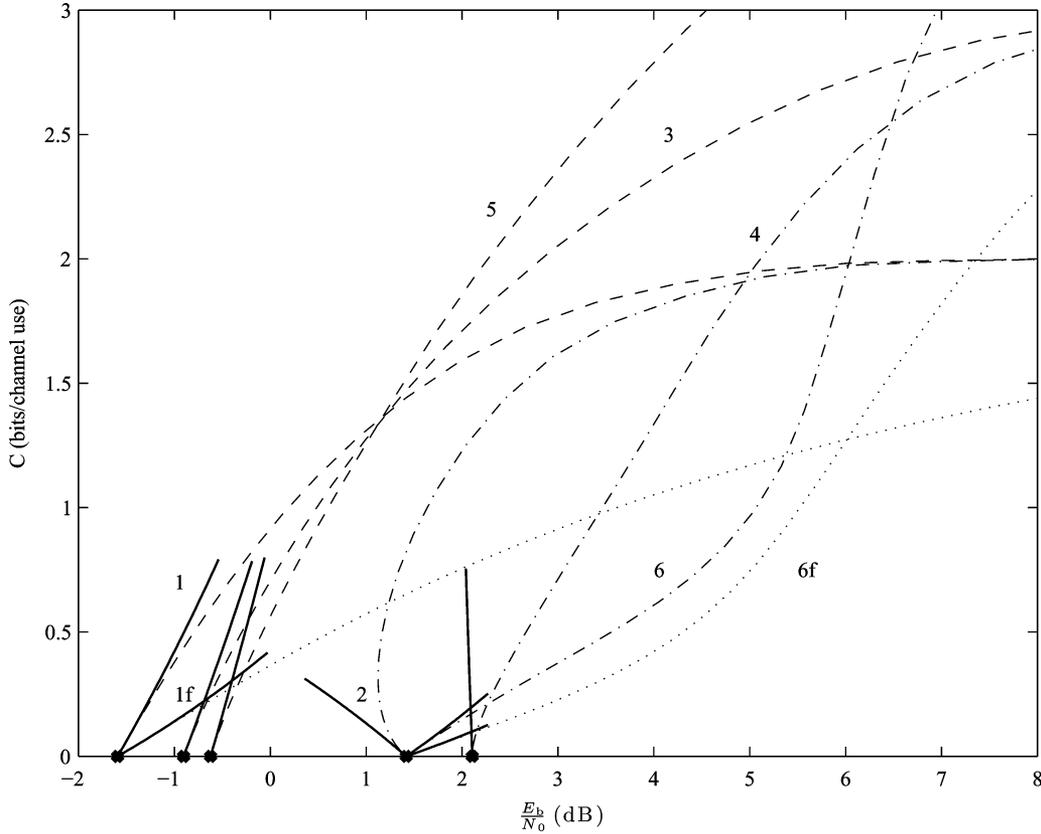
Fig. 2. BICM capacity (in bits per channel use). Labels 1 and 2 are QPSK, 3 and 4 are 8-PSK, and 5 and 6 are 16-QAM. Gray and set partitioning labeling rules correspond to dashed (and odd labels) and dashed-dotted lines (and even labels), respectively. Dotted lines are cases 1 and 6 with Nakagami-0.3 and Nakagami-1 (Rayleigh) fading (an "f" is appended to the label index). Solid lines are linear approximation around $\frac{E_b}{N_0\,\text{lim}}$.

Similarly, for CM systems with capacity $C_{\mathcal{X}}$, we have

$$C_{\mathcal{X}}W = c_1\frac{P}{N_0} + c_2\frac{P^2}{N_0^2 W} + O\left(\frac{P^{5/2}}{N_0^{5/2}W^{3/2}}\right). \qquad (20)$$

Following Verdú [3], we consider the following scenario. Let two alternative transmission systems with respective powers P_i and bandwidths $W_i, i = 1, 2$, achieve respective capacities per channel use C_i. The corresponding first- and second-order expansion coefficients are denoted by c_{11}, c_{21} for the first system, and c_{12}, c_{22} for the second. A natural comparison is to fix a power ratio $\Delta P = P_2/P_1$ and then solve for the corresponding bandwidth ratio $\Delta W = W_2/W_1$ so that the data rate is the same, that is $C_1 W_1 = C_2 W_2$. For instance, option 1 can be QPSK modulation and option 2 use of a high-order modulation with BICM.

A. An Approximation to the Tradeoff

When the capacities C_1 and C_2 can be evaluated, the exact tradeoff curve $\Delta W(\Delta P)$ can be computed. For low power, a good approximation is obtained by keeping the first two terms in the Taylor series. Under this approximation, we have the following result.

Theorem 4: In a neighborhood of $\text{SNR}_1 = 0$ the capacities in bits per second, $C_1 W_1$ and $C_2 W_2$ are equal when the expansion factors ΔP and ΔW are related as

$$\Delta W = \frac{(c_{22}\text{SNR}_1 + o(\text{SNR}_1))(\Delta P)^2}{c_{11} + c_{21}\text{SNR}_1 + o(\text{SNR}_1) - c_{12}\Delta P} \qquad (21)$$

for ΔW as a function of ΔP and, if $c_{12} \neq 0$

$$\Delta P = \frac{c_{11}}{c_{12}} + \left(\frac{c_{21}}{c_{12}} - \frac{c_{22}c_{11}^2}{c_{12}^3\Delta W}\right)\text{SNR}_1 + o(\text{SNR}_1), \quad (22)$$

for ΔP as a function of ΔW.

Proof: The proof can be found in Appendix F $\qquad \square$

Replacing the value of ΔP from (22) into (21) gives $\Delta W = \Delta W(1 + o(\text{SNR}_1))$, which is not exact, but valid within the approximation order, since we assume $\text{SNR} \to \infty$.

The previous theorem leads to the following derived results. For simplicity, we drop the terms $o(\text{SNR}_1)$ and replace the equality signs by approximate equalities.

Corollary 4: For $\Delta P = 1$, we obtain

$$\Delta W \simeq \frac{c_{22}\text{SNR}_1}{c_{11} + c_{21}\text{SNR}_1 - c_{12}}, \qquad (23)$$

and for the specific case $c_{11} = c_{12}$, $\Delta W \simeq c_{22}/c_{21}$.

The latter formula has also been obtained by Verdú [3] as a ratio of wideband slopes.

As noticed in [3], the loss in bandwidth may be significant when $\Delta P = 1$. But this point is just one of a curve relating ΔP and ΔW. For instance, with no bandwidth expansion we have the following.

Corollary 4: For $c_{11} = c_{12} = 1$, and choosing $\Delta W = 1$, $\Delta P \simeq 1 + (c_{21} - c_{22})\text{SNR}_1$.
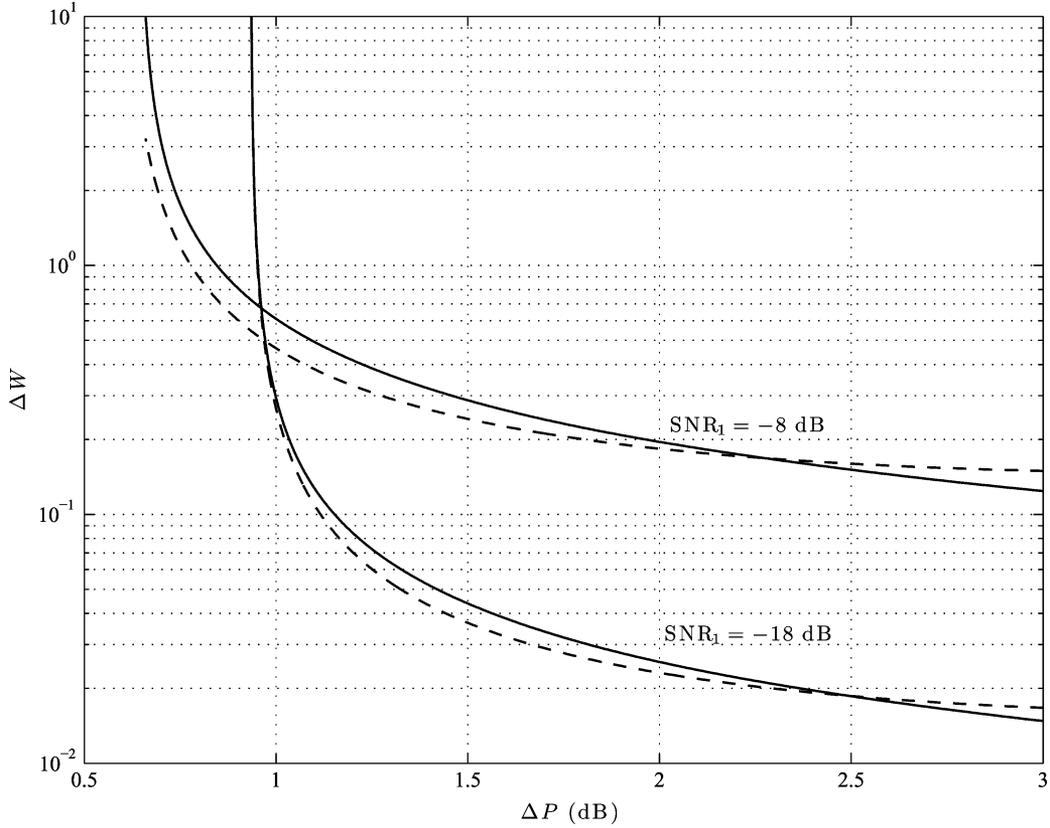
Fig. 3.   Tradeoff between ΔP and ΔW between QPSK and 16-QAM with Gray mapping. Solid lines correspond to the exact tradeoff, while dashed lines correspond to the low-SNR tradeoff.

For SNRs below -10 dB, the approximation in Theorem 4 seems accurate for "reasonable" power or bandwidth expansion ratios. A quantitative definition would lead to the problem of the extent to which the second-order approximation to the capacity is correct, a question on which we do not dwell further.

Another example concerns the effect of fully interleaved fading. Let us consider a Nakagami-ν fading model, such that the squared fading coefficient $\chi_k = |h_k|^2$ follows a gamma distribution. The parameter ν is a real positive number, $0 < \nu < \infty$. Using the values of the moments of the gamma distribution $\mathrm{E}[\chi] = 1$ and $\mathrm{E}[\chi^2] = 1 + 1/\nu$, we have that $c_1 = c_1{}^{\mathrm{AWGN}}$ and $c_2 = (1 + \frac{1}{\nu})c_2{}^{\mathrm{AWGN}}$.

Corollary 5: Consider a modulation set $\mathcal{X}$ with average unit energy and used with power P, bandwidth W, and signal-to-noise ratio SNR; its capacity in absence of fading is characterized at low SNR by the coefficients $c_1 = 1$ and c_2. When used in the Nakagami-ν channel with power P_ν and bandwidth W_ν, if $P_\nu = P$, $W_\nu = W(1 + \frac{1}{\nu})$, and if $W_\nu = W$, $P_\nu = P(1 - \frac{c_2}{\nu}\mathsf{SNR})$. As expected, for unfaded AWGN, when $\nu \to \infty$, there is no loss. Rayleigh fading ($\nu = 1$) incurs in a bandwidth expansion of a factor 2 if the power is to be fixed. On the other hand, if bandwidth is kept unchanged, there is a power penalty in decibels of about $10\log_{10}(1 - c_2\mathsf{SNR}) \simeq -10c_2\mathsf{SNR}/\log 10 \simeq -4.343 c_2 \mathsf{SNR}$ decibels, a negligible amount to all practical effects since $\mathsf{SNR} \to 0$. The worst possible fading is $\nu \to 0$, which requires an unbounded bandwidth expansion or an unlimited power penalty.

B. Tradeoff for BICM

The tradeoff between power and bandwidth can also be applied to determine the expansion factors when BICM with a nonbinary modulation is used rather than, say, QPSK modulation. Fig. 3 shows the tradeoff between power and bandwidth expansion factors when BICM over 16-QAM with Gray mapping is used, having taken QPSK as the reference transmission method. Results are presented for two values of the SNR for the QPSK baseline. The exact result, obtained by using the exact formulas for $C_{\mathcal{X}}$ and $C_{\mathcal{X},\gamma}$, respectively, (3) and (6), is plotted along the result by using Theorem 4.

As expected, for very low values of SNR, the curve for ΔW diverges as ΔP approaches the value $\frac{c_{11}}{c_{12}} = \frac{1}{0.8}$, or 0.97 dB. This is in line with the fact that the minimum energy per bit required for 16-QAM/BICM is -0.63 dB, as given in Table I. Close to this limit, small improvements in power efficiency are extremely costly in bandwidth resources. On the other hand, this loss may be accompanied by a significant reduction in bandwidth, which might be of interest in some applications. For instance, a loss of 2.4 dB from the baseline at -18 dB requires a tiny fraction of the original bandwidth, about 2%.

Concerning the last point, the results are exclusive to BICM and the same analysis can be applied to a single transmission method with coefficients c_1 and c_2, trading off power against bandwidth. In this case, for a given ΔP we would have

$$\Delta W \simeq \frac{c_2 \mathsf{SNR}_1 (\Delta P)^2}{c_1(1 - \Delta P) + c_2 \mathsf{SNR}_1}. \qquad (24)$$

Using QPSK ($c_1 = 1, c_2 = -\frac{1}{2}$) and for $\mathsf{SNR} = -18$ dB, a loss of 2.4 dB is linked to using only 3% of the original bandwidth. We see that QPSK is slightly more inefficient than BICM/16-QAM in using the bandwidth, the reason being that it has a lower coefficient c_2, -0.5 instead of -0.16. To any extent, it should not be surprising that communication in the wideband regime can be inefficient in using the bandwidth, since we are working in a regime where the main limitation is in power.

For SNRs larger than those reported in the figure, the assumption of low SNR loses its validity and the results derived from the Taylor expansion are no longer accurate.

V. CONCLUSION

In this paper, we have computed the first two derivatives of the constrained capacity at zero SNR for rather general modulation sets, and used the result to characterize analytically the behavior of BICM in the low-power regime. For binary reflected Gray mapping, the capacity loss at low SNR with respect to coded modulation is shown to be bounded by approximately 1.25 dB. This fact may be useful for the design of systems operating at low SNRs.

Moreover, we have determined the tradeoff at low SNR between power penalty and bandwidth expansion between two alternative systems. The tradeoff presented here generalizes Verdú's analysis of the wideband regime, where the bandwidth expansion for a fixed power was estimated. We have shown that no bandwidth expansion may be achieved at a negligible (but nonzero) cost in power. A similar tradeoff between power penalty and bandwidth expansion for general Nakagami-ν fading has been computed, with similar conclusions as in the point above: bandwidth expansion may be large at no power cost, but absent at a tiny power penalty. We have applied the tradeoff to a comparison between QPSK and 16-QAM.

APPENDIX A
PROOF OF PROPOSITION 1

By definition, the BICM capacity is the sum over $i = 1, \dots, m$ of the mutual informations $I(B_i; Y)$. We rewrite this mutual information as

$$I(B_i; Y) = \frac{1}{2} \sum_{b \in \{0,1\}} \mathrm{E}\left[\log \frac{\sum_{x' \in \mathcal{X}_b^i} p_{Y|X,H}(y|x', h)}{\frac{1}{2} \sum_{x' \in \mathcal{X}} p_{Y|X,H}(y|x', h)} \right] \quad (25)$$

$$= \frac{1}{2} \sum_{b \in \{0,1\}} \mathrm{E}\left[\log \left(\frac{\sum_{x' \in \mathcal{X}_b^i} \frac{2}{|\mathcal{X}|} p_{Y|X,H}(y|x', h)}{p_{Y|X,H}(y|x, h)} \right. \right.$$
$$\left. \left. \times \frac{p_{Y|X,H}(y|x, h)}{\frac{1}{2} \sum_{x' \in \mathcal{X}} \frac{2}{|\mathcal{X}|} p_{Y|X,H}(y|x', h)} \right) \right] \quad (26)$$

where we have modified the variable in the logarithm by including a factor $\frac{2}{|\mathcal{X}|} p_{Y|X,H}(y|x, h)$ in both numerator and denominator. Splitting the logarithm

$$I(B_i; Y) = \frac{1}{2} \sum_{b \in \{0,1\}} \mathrm{E}\left[\log \frac{\sum_{x' \in (\mathcal{X}_b^i)} \frac{2}{|\mathcal{X}|} p_{Y|X,H}(y|x', h)}{p_{Y|X,H}(y|x, h)} \right]$$
$$+ \frac{1}{2} \sum_{b \in \{0,1\}} \mathrm{E}\left[\log \frac{p_{Y|X,H}(y|x, h)}{\frac{1}{|\mathcal{X}|} \sum_{x' \in \mathcal{X}} p_{Y|X,H}(y|x', h)} \right]. \quad (27)$$

For fixed b, the quantity

$$-\mathrm{E}\left[\log \frac{\sum_{x' \in \mathcal{X}_b^i} \frac{2}{|\mathcal{X}|} p_{Y|X,H}(y|x', h)}{p_{Y|X,H}(y|x, h)} \right] \quad (28)$$

is the mutual information achievable by using equiprobable signaling in the set $\mathcal{X}_b^i$, $\mathrm{C}_{\mathcal{X}_b^i}^{\mathrm{u}}$, and, similarly, the quantity

$$\mathrm{E}\left[\log \frac{p_{Y|X,H}(y|x, h)}{\frac{1}{|\mathcal{X}|} \sum_{x' \in \mathcal{X}} p_{Y|X,H}(y|x', h)} \right] \quad (29)$$

is the mutual information achieved by equiprobable signaling in $\mathcal{X}$, $\mathrm{C}_{\mathcal{X}}^{\mathrm{u}}$.

APPENDIX B
LINEAR EXPANSION CAPACITY

We start with (7) and use Lagrange's inversion formula. The inversion formula transforms a function

$$\mathrm{C} = f_1(\mathsf{SNR}) \quad (30)$$

into its inverse

$$\mathsf{SNR} = f_2(\mathrm{C}). \quad (31)$$

We do an expansion around $\mathsf{SNR} = 0$, which is also $\mathrm{C} = 0$. Applied to our case, the inversion formula becomes

$$\mathsf{SNR} = \frac{\mathsf{SNR}}{f_1(\mathsf{SNR})} \bigg|_{\mathsf{SNR} \to 0} \mathrm{C}$$
$$+ \frac{1}{2} \frac{d}{d\mathsf{SNR}} \left(\frac{\mathsf{SNR}}{f_1(\mathsf{SNR})} \right)^2 \bigg|_{\mathsf{SNR} \to 0} \mathrm{C}^2 + \mathrm{O}(\mathrm{C}^3). \quad (32)$$

Using the expansion in (7), after some simplifications we get

$$\mathsf{SNR} = \frac{\log 2}{c_1} \mathrm{C} - \frac{c_2 \log^2 2}{c_1^3} \mathrm{C}^2 + \mathrm{O}(\mathrm{C}^3). \quad (33)$$

Letting $\mathsf{SNR} = \mathrm{C} \frac{E_{\mathrm{b}}}{N_0}$ and rearranging we obtain

$$\frac{E_{\mathrm{b}}}{N_0} = \frac{\log 2}{c_1} - \frac{c_2 \log^2 2}{c_1^3} \mathrm{C} + \mathrm{O}(\mathrm{C}^2) \quad (34)$$

which leads to

$$\mathrm{C} = -\frac{\log^2 2 c_2}{c_1^3} \left(\frac{E_{\mathrm{b}}}{N_0} - \frac{\log 2}{c_1} \right) + \mathrm{O}\left(\left(\frac{E_{\mathrm{b}}}{N_0} - \frac{\log 2}{c_1} \right)^2 \right)$$

and hence the desired result.

APPENDIX C
CM CAPACITY EXPANSION AT LOW SNR

The assumption that the constellation moments are finite implies that $\mathrm{E}[|X|^{2+\alpha}] < \infty$ for $\alpha > 0$. Therefore, as $\mathsf{SNR} \to 0$, for $\mu > 0$, the technical condition

$$\mathsf{SNR}^{2+\alpha} \mathrm{E}[|X|^{2+\alpha}] \leq (-\log \sqrt{\mathsf{SNR}})^{\mu}, \quad (35)$$

necessary to apply Theorem 5 of [7] holds.

Let us define a 2×1 vector $x^{(r)} = (x_r \; x_i)^T$, with components the real and imaginary parts of symbol x, respectively, denoted by x_r and x_i. The covariance matrix of $x^{(r)}$, denoted by $\mathrm{cov}(X)$, is given by

$$
\mathrm{cov}(X) \\
= \begin{pmatrix} \mathrm{E}[(Xr - \hat{x}_r)^2] & \mathrm{E}[(X_r - \hat{x}_r)(X_i - \hat{x}_i)] \\ \mathrm{E}[(X_r - \hat{x}_r)(X_i - \hat{x}_i)] & \mathrm{E}[(X_i - \hat{x}_i)^2] \end{pmatrix} \tag{36}
$$

where $\hat{x}_r$ and $\hat{x}_i$ are the mean values of the real and imaginary parts of the constellation.

Theorem 5 of [7] gives $c_1 = \mathrm{Tr}(\mathrm{cov}(X))$ and $c_2 = -\mathrm{Tr}(\mathrm{cov}^2(X))$, or

$$
c_1 = \mathrm{E}[(X_r - \hat{x}_r)^2] + \mathrm{E}[(X_i - \hat{x}_i)^2] \tag{37}
$$
$$
c_2 = -(\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \mathrm{E}^2[(X_i - \hat{x}_i)^2] \\ + 2\mathrm{E}^2[(X_r - \hat{x}_r)(X_i - \hat{x}_i)]). \tag{38}
$$

The coefficient c_1 coincides with that in (12).

As for c_2, let us add a subtract a term $\mathrm{E}[(X_r - \hat{x}_r)^2]\mathrm{E}[(X_i - \hat{x}_i)^2]$ to (38). Then

$$
c_2 = -\Big(\tfrac{1}{2}\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \tfrac{1}{2}\mathrm{E}^2[(X_i - \hat{x}_i)^2] \\
+ \mathrm{E}[(X_r - \hat{x}_r)^2]\mathrm{E}[(X_i - \hat{x}_i)^2] \\
+ \tfrac{1}{2}\mathrm{E}^2[(X_r - \hat{x}_r)^2] + \tfrac{1}{2}\mathrm{E}^2[(X_i - \hat{x}_i)^2] \\
- \mathrm{E}[(X_r - \hat{x}_r)^2]\mathrm{E}[(X_i - \hat{x}_i)^2] \\
+ 2\mathrm{E}^2[(X_r - \hat{x}_r)(X_i - \hat{x}_i)] \Big) \tag{39}
$$

which, in turn, can be written as

$$
c_2 = -\tfrac{1}{2}(\mathrm{E}^2[|X - \hat{x}|^2] + |\mathrm{E}[(X - \hat{x})^2]|^2), \tag{40}
$$

a form which coincides with (13), by noting that

$$
\mathrm{E}[|X - \hat{x}|^2] = \mathrm{E}[|X|^2] - |\hat{x}|^2 = \mu_2(\mathcal{X}) - |\mu_1(\mathcal{X})|^2 \tag{41}
$$
$$
\mathrm{E}[(X - \hat{x})^2] = \mathrm{E}[X^2] - \hat{x}^2 = \mu_2'(\mathcal{X}) - \mu_1^2(\mathcal{X}). \tag{42}
$$

APPENDIX D
PROOF OF THEOREM 2

In (6) for the BICM capacity, the summands $C_{\mathcal{X}}$ and $C_{\mathcal{X}_b^i}$ admit each a Taylor expansion given in Theorem 1. Hence

$$
c_1 = \sum_{i=1}^m \frac{1}{2} \sum_{b=0,1} \left(1 - \left(\mu_2(\mathcal{X}_b^i) - |\mu_1(\mathcal{X}_b^i)|^2 \right) \right) \tag{43}
$$
$$
= \sum_{i=1}^m \left(\left(1 - \frac{1}{2}\sum_{b=0,1} \mu_2(\mathcal{X}_b^i) \right) + \frac{1}{2}\sum_{b=0,1} |\mu_1(\mathcal{X}_b^i)|^2 \right) \tag{44}
$$
$$
= \sum_{i=1}^m \left(\left(\sum_{s\in\mathcal{X}} \frac{1}{|\mathcal{X}|}|s|^2 - \sum_{s\in\mathcal{X}} \frac{1}{|\mathcal{X}|}|s|^2 \right) \right. \\
\left. + \frac{1}{2}\sum_{b=0,1}\sum_{s\in(\mathcal{X}_b^i)} \frac{2}{|\mathcal{X}|} |\mu_1(\mathcal{X}_b^i)|^2 \right) \tag{45}
$$

$$
= \sum_{i=1}^m \frac{1}{2} \sum_{b=0,1} |\mu_1(\mathcal{X}_b^i)|^2 \tag{46}
$$

since $\frac{1}{2}\sum_{b=0,1}\mu_2(\mathcal{X}_b^i) = \mu_2(\mathcal{X}) = 1$ by construction.

As for c_2, it follows from a similar application of Theorem 1.

APPENDIX E
FIRST-ORDER COEFFICIENT FOR BICM WITH GRAY MAPPING

For M-PAM, the Gray mapping construction makes $\mu_1(\mathcal{X}_b^i) = 0$, for $b = 0, 1$ and all bit positions except one, which we take with no loss of generality to be $i = 1$. Therefore

$$
c_1 = \frac{1}{2}|\mu_1(\mathcal{X}_1^0)|^2 + \frac{1}{2}|\mu_1(\mathcal{X}_1^1)|^2 = |\mu_1(\mathcal{X}_1^0)|^2 = |\mu_1(\mathcal{X}_1^1)|^2. \tag{47}
$$

The last equalities follow from the symmetry between 0 and 1.

Symbols lie on a line in the complex plane with values $\pm\beta(1, 3, 5, \ldots, M - 1)$, with β a normalization factor $\beta^2 = 3/(M^2 - 1)$. This factor follows by setting $2n = M$ in the formula $\frac{1}{n}\sum_{i=1}^n (2i-1)^2 = \frac{1}{3}((2n)^2 - 1)$, The average symbol has modulus $|\mu_1(\mathcal{X}_1^0)| = \beta\frac{M}{2}$, and therefore

$$
c_1 = |\mu_1(\mathcal{X}_1^0)|^2 = \frac{3 \cdot M^2}{4(M^2 - 1)}. \tag{48}
$$

Extension to M^2-QAM is clear, by taking the Cartesian product along real and imaginary parts. Now, two indices i contribute, each with an identical form to that of pulse amplitude modulation (PAM). As the energy along each axis of half that of PAM, the normalization factor β_{QAM}^2 also halves and overall c_1 does not change.

APPENDIX F
DETERMINATION OF THE POWER AND BANDWIDTH TRADE-OFF

In order to have the same capacities, bandwidth and/or power must change to account for the difference in capacity, so that

$$
c_{11}\frac{P_1}{N_0} + c_{21}\frac{P_1^2}{W_1 N_0^2} + \mathrm{o}(W_1 \mathrm{SNR}_1^2) \\
= c_{12}\frac{P_2}{N_0} + c_{22}\frac{P_2^2}{W_2 N_0^2} + \mathrm{o}(W_2 \mathrm{SNR}_2^2). \tag{49}
$$

Simplifying common factors, we obtain

$$
c_{11} + c_{21}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) \\
= c_{12}\frac{P_2}{P_1} + (c_{22} + \mathrm{o}(\mathrm{SNR}_1))\frac{P_2^2}{P_1^2}\frac{W_1}{W_2}\mathrm{SNR}_1. \tag{50}
$$

Or, with the definitions $\Delta P = P_2/P_1$, and $\Delta W = W_2/W_1$

$$
c_{11} + c_{21}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) \\
= c_{12}\Delta P + (c_{22}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1))\frac{(\Delta P)^2}{\Delta W} \tag{51}
$$

and

$$
\Delta W = \frac{(c_{22}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1))(\Delta P)^2}{c_{11} + c_{21}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1) - c_{12}\Delta P}. \tag{52}
$$

This equation gives the tradeoff between ΔP and ΔW, for a fixed (small) SNR_1, so that the capacities of scenarios 1 and 2 coincide.

Next we solve for the inverse, i.e., for ΔP as a function of ΔP. First, let us define the quantities $a = c_{22}\mathrm{SNR}_1 + \mathrm{o}(\mathrm{SNR}_1)$

and $b = c_{11} + c_{21}\mathsf{SNR}_1 + \mathrm{o}(\mathsf{SNR}_1)$. Then, rearranging (52), we have $a(\Delta P)^2 + c_{12}\Delta W \Delta P - b\Delta W = 0$ and therefore

$$\Delta P = \frac{-c_{12}\Delta W \pm \sqrt{(c_{12}\Delta W)^2 + 4ab\Delta W}}{2a} \qquad (53)$$

$$= \frac{c_{12}\Delta W}{2a}\left(-1 \pm \sqrt{1 + \frac{4ab}{c_{12}{}^2\Delta W}}\right). \qquad (54)$$

Often we have $c_{22} < 0$, and then the negative root is a spurious solution. We choose then the positive root. Since ab is of order SNR_1, we can use the Taylor expansion $(1 + 4t)^{1/2} = 1 + 2t - 2t^2 + \mathrm{o}(t^2)$, to write

$$\Delta P = \frac{c_{12}\Delta W}{2a}\left(\frac{2ab}{c_{12}{}^2\Delta W} - \frac{2a^2b^2}{c_{12}{}^4(\Delta W)^2}\right) \qquad (55)$$

$$= \frac{b}{c_{12}} - \frac{ab^2}{c_{12}{}^3\Delta W}. \qquad (56)$$

Since $\mathsf{SNR}_1 \to 0$, we group the nonlinear terms in SNR_1 and so get

$$\Delta P = \frac{c_{11} + c_{21}\mathsf{SNR}_1}{c_{12}} - \frac{c_{22}c_{11}{}^2\mathsf{SNR}_1}{c_{12}{}^3\Delta W} + \mathrm{o}(\mathsf{SNR}_1) \qquad (57)$$

$$= \frac{c_{11}}{c_{12}} + \left(\frac{c_{21}}{c_{12}} - \frac{c_{22}c_{11}{}^2}{c_{12}{}^3\Delta W}\right)\mathsf{SNR}_1 + \mathrm{o}(\mathsf{SNR}_1). \qquad (58)$$

REFERENCES

[1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, May 1992.

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

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

[4] G. D. Forney, Jr. and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2384–2415, Oct. 1998.

[5] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[6] F. Brännström and L. K. Rasmussen, "Classification of 8 PSK mappings for BICM," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 2136–2140.

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

[8] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.