

# Probabilistic Signal Shaping for Bit-Metric Decoding

Georg Böcherer

Institute for Communications Engineering  
Technische Universität München, Germany  
Email: georg.boecherer@tum.de

**Abstract**—A scheme is proposed that combines probabilistic signal shaping with bit-metric decoding. The transmitter generates symbols according to a distribution on the channel input alphabet. The symbols are labeled by bit strings. At the receiver, the channel output is decoded with respect to a bit-metric. An achievable rate is derived using random coding arguments. For the 8-ASK AWGN channel, numerical results show that at a spectral efficiency of 2 bits/s/Hz, the new scheme outperforms bit-interleaved coded modulation (BICM) without shaping and BICM with bit shaping (i Fabregas and Martinez, 2010) by 0.87 dB and 0.15 dB, respectively, and is within 0.0094 dB of the coded modulation capacity. The new scheme is implemented by combining a distribution matcher with a systematic binary low-density parity-check code. The measured finite-length gains are very close to the gains predicted by the asymptotic theory.

## I. INTRODUCTION

*Bit-interleaved coded modulation* (BICM) combines high order modulation with *binary* error correcting codes [1]. This makes BICM attractive for practical application and BICM is widely used in standards, e.g., in DVB-T2/S2/C2. At a BICM receiver, *bit-metric decoding* is used [2, Sec. II]. We are *not* considering BICM with *iterative demapping-decoding* (BICM-ID, [3]) here.

For bit-metric decoding, the signal points of a channel input constellation of size  $2^m$  are labeled by bit strings of length  $m$ . The  $m$  bit levels are treated independently at the decoder. We make this precise in Sec. II. Let  $\mathbf{B} = (B_1, B_2, \dots, B_m)^T$  denote a column vector of  $m$  binary random variables  $B_i$ ,  $i = 1, 2, \dots, m$ , representing the bit levels. Conditioned on the channel input represented by  $\mathbf{B}$ , let  $P_{Y|\mathbf{B}}$  be the distribution of the channel output  $Y$ . In [4], it was shown that a bit-metric decoder achieves the rate

$$\sum_{i=1}^m \mathbb{I}(B_i; Y) \quad (1)$$

where  $\mathbb{I}(B_i; Y)$  denotes the mutual information of bit level  $B_i$  and channel output  $Y$ . The proof in [4] assumes an ideal interleaver, which lets the  $m$  bit levels see  $m$  independent binary input channels. The authors in [2] showed that (1) is achievable without an ideal interleaver for independent and uniformly distributed  $B_i$ . We call this scheme *uniform BICM*. In [5], the authors showed that (1) is achievable for independent and arbitrarily distributed bit levels  $B_i$ . We call this scheme *bit shaped BICM* (BS-BICM). It was illustrated

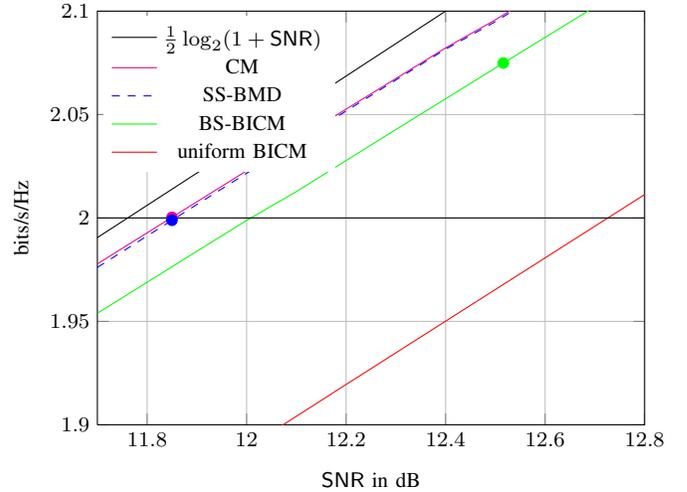


Fig. 1. Power-rate curves for the 8-ASK AWGN channel. The points on the curves that are marked by dots are discussed in Sec. III-E.

in [5] for the *additive white Gaussian noise* (AWGN) channel that by properly choosing the bit level distributions, a shaping gain over uniform BICM can be achieved. The authors of [6] proposed an algorithm to calculate the maximum rate of bit shaped BICM.

In Sec. II, we show that the rate

$$\left[ \sum_{i=1}^m \mathbb{I}(B_i; Y) \right] - \left\{ \left[ \sum_{i=1}^m \mathbb{H}(B_i) \right] - \mathbb{H}(\mathbf{B}) \right\} \quad (2)$$

is achievable by a bit-metric decoder and for any *joint* input distribution  $P_{\mathbf{B}}$ .  $\mathbb{H}$  denotes the entropy function. We call our scheme *signal shaping with bit-metric decoding* (SS-BMD). In Sec. III, we calculate the power-rate curves for the AWGN channel with an equidistant 8-ASK input constellation. The resulting power-rate curves are displayed in Fig. 1. We find that at a spectral efficiency of 2 bits/s/Hz, SS-BMD outperforms uniform BICM and BS-BICM by 0.87 dB and 0.15 dB, respectively, and it lies within 0.0094 dB of the *coded modulation* (CM) capacity. We then present in Sec. V implementations of SS-BMD and BS-BICM where we combine a distribution matcher with the DVB-S2 rate 3/4 code [7]. To the best of our knowledge, no other implementation of BS-BICM has been reported in literature so far. *We emphasize that we are*

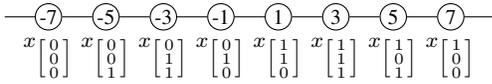


Fig. 2. The *Binary Reflected Gray Code* as defined in [12, Sec. II.B].

considering probabilistic shaping for BICM without iterative demapping-decoding. Probabilistic shaping for BICM-ID has been considered in [8], constellation shaping for BICM in [9] and constellation shaping for BICM-ID in [10]. We compare to an implementation of uniform BICM with the DVB-S2 rate 2/3 code. At 2 bits/s/Hz and end-to-end block error probability of  $10^{-2}$ , SS-BMD outperforms uniform BICM and BS-BICM by 0.83 dB and 0.13 dB, respectively. SS-BMD operates within 0.9 dB of the CM capacity and 0.99 dB of the Shannon capacity. In [11], similar results were achieved by combining probabilistic shaping with multilevel coding and multi stage decoding.

## II. ACHIEVABLE RATE FOR SS-BMD

Let  $P_{Y|B}$  be a *discrete memoryless channel* (DMC) with input  $\mathbf{B} = (B_1, B_2, \dots, B_m)^T$  and output  $Y$ . The  $B_i$  are binary random variables. A *bit-metric decoder* uses the metric

$$q(\mathbf{b}, y) = \prod_{i=1}^m q_i(b_i, y) \quad (3)$$

where  $q_i(b_i, y)$  is a function of

$$P_{B_i Y}(b_i, y) = \sum_{\mathbf{a} \in \{0,1\}^m : a_i = b_i} P_{Y|\mathbf{B}}(y|\mathbf{a}) P_{\mathbf{B}}(\mathbf{a}). \quad (4)$$

If the  $B_i$  are independent, then our definition is equivalent to [5, Eq. (9)]. If the  $B_i$  are also uniformly distributed, then our definition is equivalent to [2, Eq. (6)].

**Theorem 1.** For a DMC  $P_{Y|B}$  with finite input and output alphabets, the rate (2) is achievable by a bit-metric decoder.

*Proof:* The proof is given in the appendix. ■

## III. $2^m$ -ASK MODULATION FOR THE AWGN CHANNEL

The discrete time baseband AWGN channel is described by

$$Y = X + Z \quad (5)$$

where  $X$  and  $Y$  are the input and output, respectively, and where  $Z$  is zero mean and unit variance Gaussian noise. If the input is subject to an average power constraint SNR, the capacity of the AWGN channel is [13, Theo. 7.4.2]

$$C(\text{SNR}) = \frac{1}{2} \log_2(1 + \text{SNR}). \quad (6)$$

In practice, the input is restricted to a finite set  $\mathcal{X}$  of constellation points. We consider *amplitude shift keying* (ASK) with  $2^m$  equidistant constellation points, i.e., we have

$$\mathcal{X}_{\text{ASK}} = \{2 \cdot i - 2^m - 1 : i = 1, 2, \dots, 2^m\}. \quad (7)$$

The points  $x \in \mathcal{X}_{\text{ASK}}$  are labeled by a binary vector  $\mathbf{B} = (B_1, \dots, B_m)^T$ . See Fig. 2 for an example with  $m = 3$ .

We model the channel input as  $d \cdot x_{\mathbf{B}}$  where the label  $\mathbf{B}$  is distributed according to the distribution  $P_{\mathbf{B}}$  and where  $d$  is a non-negative real number that scales the constellation.

### A. Coded modulation

The *coded modulation* (CM) capacity of the ASK constellation is

$$C_{\text{CM}}(\text{SNR}) = \max_{P_{\mathbf{B}}, d} \mathbb{I}(\mathbf{B}; Y) \quad (8)$$

subject to  $d^2 \mathbb{E}[x_{\mathbf{B}}^2] \leq \text{SNR}$ .

The maximization is both over the distribution  $P_{\mathbf{B}}$  and the constellation scaling  $d$ .

### B. Bit Shaped BICM

In BICM, the labeling of the constellation strongly influences the achievable rate, see [12]. We denote by  $L$  the chosen labeling. An achievable rate for bit shaped BICM is by [5, Eq. (19)]

$$C_{\text{BICM}}^L(\text{SNR}) = \max_{P_{\mathbf{B}}, d} \sum_{i=1}^m \mathbb{I}(B_i; Y) \quad (9)$$

subject to  $P_{\mathbf{B}} = \prod_{i=1}^m P_{B_i}$

$d^2 \mathbb{E}[x_{\mathbf{B}}^2] \leq \text{SNR}$

The optimization is over the constellation scaling  $d$  and the bit distributions  $P_{B_i}$ . When all bit distributions  $P_{B_i}$  are uniform, we get the achievable rate of uniform BICM.

### C. SS-BMD

**Theorem 2.** SS-BMD achieves the rate

$$C_{\text{SS-BMD}}^L(\text{SNR}) = \max_{P_{\mathbf{B}}, d} \sum_{i=1}^m \mathbb{I}(B_i; Y) - \left[ \sum_{i=1}^m \mathbb{H}(B_i) - \mathbb{H}(\mathbf{B}) \right] \quad (10)$$

subject to  $d^2 \mathbb{E}[x_{\mathbf{B}}^2] \leq \text{SNR}$ .

*Proof:* The theorem can be proven by adapting the proof of Theo. 1 to the AWGN channel, similar to the approach taken in [2, Sec. III]. ■

### D. Numerical Results

We evaluate achievable rates for 8-ASK near a spectral efficiency of 2 bits/s/Hz. We choose the Gray labeling in Fig. 2, since it is the best known labeling near 2 bits/s/Hz for BICM [12, Fig. 2(b)]. We discretize the channel output  $y$  into  $2^9 = 512$  intervals; this choice effectively achieves the continuous CM capacity. We calculate the CM capacity by line search over  $d$ ; for each value of  $d$ , maximization over  $P_{\mathbf{B}}$  is a convex optimization problem. We use CVX [14] to solve the problem. For bit shaped BICM, we again do a line search over  $d$ . For each value of  $d$ , maximization over  $P_{B_1} P_{B_2} P_{B_3}$  is a non-convex optimization problem [6]. We use the algorithm from [6], [15] to solve the problem. Uniform BICM is straightforward. For SS-BMD, we use a heuristic. We evaluate the SS-BMD achievable rate in the values  $d, P_{\mathbf{B}}$  that achieve the CM capacity. In Fig. 1 we display the resulting power-rate curves.

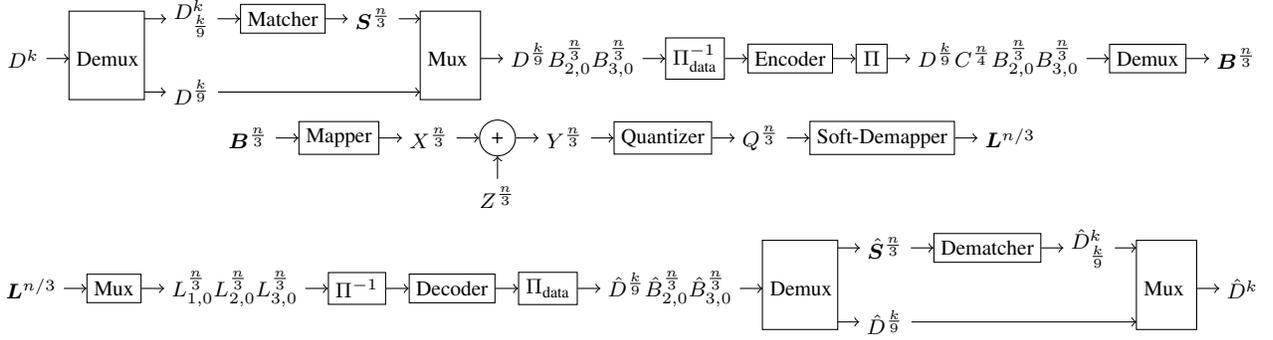


Fig. 3. Transmitter and receiver of the SS-BMD system. *Mux* is short for *multiplexer* and *Demux* is short for *demultiplexer*.

### E. Discussion

The CM capacity of 8-ASK is an upper bound for any 8-ASK transmission scheme. At 2 bits/s/Hz, SS-BMD gets within 0.0094 dB of CM capacity while for BS-BICM, the gap to CM capacity is 0.16 dB. This is somewhat surprising, since we would expect that SS-BMD loses in terms of *rate* over BS-BICM because of the correlated bit levels. In fact, this does happen, but SS-BMD gains in terms of *SNR*. The dots in Fig. 1 illustrate this. The blue and the red dot are obtained by evaluating the CM capacity and the SS-BMD rate in the same distribution  $P_{\mathcal{B}}$ . The green dot is obtained by evaluating the BS-BICM rate in the marginals  $P_{B_1}, P_{B_2}, P_{B_3}$  of  $P_{\mathcal{B}}$ . Using the marginals leads to a rate gain of 0.076 bits per channel use compared to SS-BMD, however, it also leads to an SNR loss of 0.67 dB, which moves it away from the CM capacity curve. For all three dots, we used the same constellation scaling  $d$ .

### IV. SS-BMD SYSTEM DESIGN

We design a system that lets us reap the shaping gap between SS-BMD and uniform BICM. The distribution  $P_{\mathcal{B}}$  that achieves CM capacity is symmetric around zero. The first bit level of Gray labeling chooses the sign of the constellation point, see Fig. 2. The labeling of the other bit levels is symmetric around zero. This means that  $B_1$  is stochastically independent of  $B_2 \cdots B_m$ , i.e., we have

$$P_{\mathcal{B}}(\mathbf{b}) = P_{B_1}(b_1)P_{B_2 \cdots B_m}(b_2 \cdots b_m), \forall \mathbf{b} \in \{0, 1\}^m. \quad (11)$$

Furthermore,  $B_1$  is uniformly distributed. Systematic binary encoders copy data bits to the codeword and append parity bits. Thus, if the data bits have a non-uniform distribution, *this distribution is preserved by the encoder*. However, the parity bits are modulo 2 sums of data bits, so their distribution is approximately uniform and it is reasonable to model the parity bits to be independent and uniformly distributed [16, Chap. 7], [17]. An SS-BMD system could mimic the capacity-achieving distribution as follows. Use a distribution matcher on data bits to generate  $B_2 \cdots B_m$  according to  $P_{B_2 \cdots B_m}$ . Encode  $B_2 \cdots B_m$  by a systematic encoder and use the parity bits and possibly additional data bits for bit level 1. This is possible as

long as the coding rate of the code fulfills

$$\frac{k}{n} \geq \frac{m-1}{m}. \quad (12)$$

Ungerböck [18] made the observation that for reliable transmission over the AWGN channel, it suffices to add one bit of redundancy per real dimension. This observation was analytically confirmed in [19] and it was experimentally confirmed in [11], for example. This suggests that condition (12) is feasible.

### V. IMPLEMENTATION

We now implement the SS-BMD system that we outlined in Sec. IV. Our target is to transmit 2 bits per channel use reliably over the AWGN channel with the Gray labeled 8-ASK input constellation, as described in Sec. III-D. We use the DVB-S2 LDPC codes [7] with block length  $n = 64800$  and code rate  $3/4$  for SS-BMD and BS-BICM, and with code rate  $2/3$  for uniform BICM. We next discuss the important parts of our SS-BMD system. A complete flow chart is provided in Fig. 3.

#### A. Notation

For row vectors, we use the notation

$$V_i^j = (V_{i+1}, V_{i+2}, \dots, V_j). \quad (13)$$

If it is clear from the context, we write  $V^j$  instead of  $V_0^j$ . The vector  $V_i^j$  thus has  $j-i$  entries and we can write  $V^j = (V_0^i, V_i^j)$  for  $0 < i < j$ .

We denote column vectors using a bold font. The labels of  $n$  consecutive ASK-8 signal points are denoted by

$$\mathbf{B}^n = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_n), \quad \mathbf{B}_i = \begin{bmatrix} B_{1,i} \\ B_{2,i} \\ B_{3,i} \end{bmatrix}, i = 1, 2, \dots, n.$$

Bit levels 2 and 3 together are denoted by  $\mathcal{S}$ , i.e., we have

$$\mathbf{S} = \begin{bmatrix} B_2 \\ B_3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} B_1 \\ \mathbf{S} \end{bmatrix}, \quad \mathbf{B}^n = \begin{bmatrix} \mathbf{B}_{1,0}^n \\ \mathbf{S}^n \end{bmatrix}. \quad (14)$$

### B. Adapt the Signal Point Distribution to the Code Rate

For the Gray labeling in Fig. 2, the distribution that achieves the SS-BMD rate of 2 bits/s/Hz is  $P_B^* = P_{B_1}^* P_S^*$  with  $P_{B_1}^*(0) = P_{B_1}^*(1) = 1/2$  and

$$P_S^*(00) = 0.0579 \quad (15)$$

$$P_S^*(01) = 0.1507 \quad (16)$$

$$P_S^*(11) = 0.3237 \quad (17)$$

$$P_S^*(10) = 0.4676. \quad (18)$$

We use a rate 3/4 code, so according to our outline in Sec. IV, we use “matched data” with joint distribution  $P_S$  on bit levels 2 and 3. We use 1/4 of bit level 1 for uniformly distributed data, and we use the remaining 3/4 of bit level 1 for parity bits. This bit level assignment is visualized in Fig. 4. The distribution  $P_{B_1}^* P_S^*$  results in

$$\frac{1}{4} \mathbb{H}(P_{B_1}^*) + \mathbb{H}(P_S^*) = 0.25 + 1.6891 = 1.9391. \quad (19)$$

We therefore need to choose a distribution for  $S$  that is close to  $P_S^*$ , but whose entropy is equal to 1.75. We choose

$$P_S = \underset{P}{\operatorname{argmin}} \mathbb{D}(P \| P_S^*)$$

subject to  $\mathbb{H}(P) \geq 1.75$  (20)

where  $\mathbb{D}(P \| P_S^*)$  denotes the *informational divergence* or *relative entropy* of  $P$  and  $P_S^*$  [20, Sec. 2.3]. The optimization problem (20) is convex and by solving the KKT conditions [21, Sec. 5.5.3], we find the solution

$$P_S(\mathbf{a}) = \frac{P_S^*(\mathbf{a})^\lambda}{\sum_{\mathbf{b} \in \{0,1\}^2} P_S^*(\mathbf{b})^\lambda} \quad (21)$$

where  $\lambda$  is chosen such that  $\mathbb{H}(P_S) = 1.75$ . We find  $\lambda = 0.8672$  and

$$P_S(00) = 0.0722 \quad (22)$$

$$P_S(01) = 0.1654 \quad (23)$$

$$P_S(11) = 0.3209 \quad (24)$$

$$P_S(10) = 0.4415. \quad (25)$$

Since we have fixed the distribution to  $P_S$ , the SNR and the achievable rate depend only on the constellation scaling  $d$ . We choose  $d$  such the SS-BMD rate evaluates to 2 bits/s/Hz. We observe a loss of 0.028 dB as compared to  $P_S^*$ .

### C. Matcher Input and Output Lengths

At the system input, we have  $k = \frac{2n}{3} = 43200$  data bits  $D^k$  that are independent and uniformly distributed. For bit level 1, we use  $\frac{n}{12} = \frac{k}{9} = 5400$  data bits  $D^{\frac{k}{9}}$  according to Fig. 4. For bit levels 2 and 3, we map the remaining  $8k/9 = 37800$  data bits  $D^{\frac{k}{9}}$  to  $\frac{n}{3} = 21600$  symbols that are distributed according to  $P_S$ .

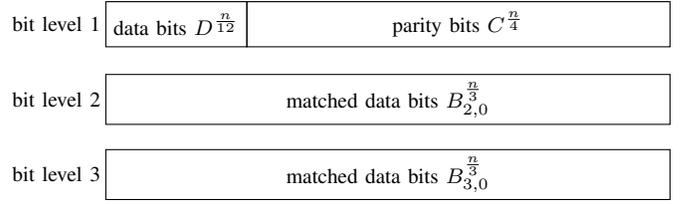


Fig. 4. Visualization of the bit levels.

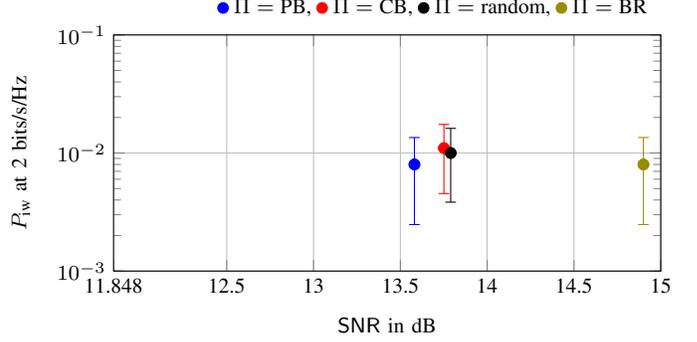


Fig. 5. SNR versus  $P_{\text{ew}} = \Pr\{D^k \neq \hat{D}^k\}$  at 2 bits/s/Hz for uniform BICM. The DVB-S2 rate 2/3 LDPC code is used with the different interleavers discussed in Sec. V-D. 11.848 dB corresponds to the CM capacity.

We verify that this approach is in accordance with the distribution  $P_S$ . An ideal distribution matcher performs a one-to-one mapping from the input to the output. Thus, information is conserved and we have

$$\frac{n}{3} \cdot \mathbb{H}(P_S) = 21600 \cdot 1.75 = 37800 = \frac{8k}{9}. \quad (26)$$

### D. Interleaver

Since the DVB-S2 codes are systematic and highly structured [7], the performance may depend on which coded bits are used for which bit level. This is controlled by the interleaver  $\Pi$ , see Fig. 3. Since we want to compare to uniform BICM, we choose an interleaver that works well for uniform BICM with the DVB-S2 rate 2/3 code. We adjust the SNR such that the information word error probability

$$P_{\text{iw}} = \Pr\{D^k \neq \hat{D}^k\} \quad (27)$$

is around  $P_{\text{iw}} = 10^{-2}$ . We try a random interleaver, the *consecutive-bit* (CB) interleaver [22], the *bit-reliability* (BR) interleaver [22], and an interleaver that is defined by the following permutation of the codeword  $V^n$ :

$$B_{1,0}^{\frac{n}{3}} = V_{43200}^{64800} \quad (28)$$

$$B_{2,0}^{\frac{n}{3}} = V_{21600}^{43200} \quad (29)$$

$$B_{3,0}^{\frac{n}{3}} = V_0^{21600}. \quad (30)$$

For the DVB-S2 codes, the parity bits are appended to the data bits. The interleaver defined in (28)–(30) copies the parity bits to bit level 1. We therefore call it the *parity-bit* (PB) interleaver. The simulation results for  $P_{\text{iw}} \approx 10^{-2}$  are

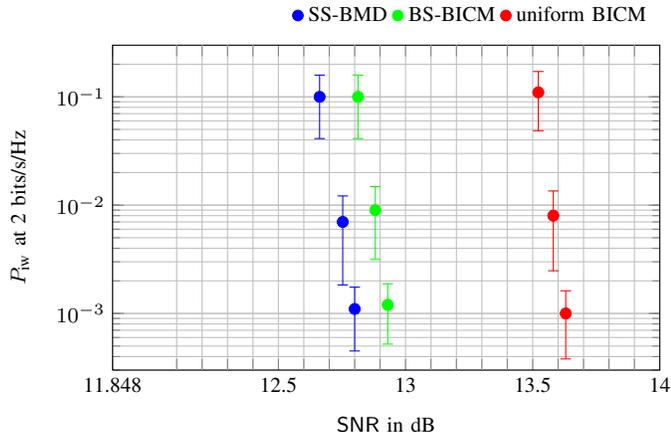


Fig. 6. SNR versus  $P_w = \Pr\{D^k \neq \hat{D}^k\}$  at 2 bits/s/Hz. The PB interleaver (see Sec. V-D) is used. For uniform BICM, the DVB-S2 rate 2/3 LDPC code is employed. For SS-BMD and BS-BICM, an arithmetic distribution matcher [23] and the DVB-S2 rate 3/4 LDPC code is used. 11.848 dB corresponds to the CM capacity.

shown in Fig. 5. The PB interleaver performs best, and we therefore use it in our implementation. Note that we made this choice based on simulation results for a specific LDPC code, namely the rate 2/3 DVB-S2 code. We do not know if the PB interleaver is the optimal choice and other codes may lead to other choices.

We use the PB interleaver also for BS-BICM and SS-BMD. To preserve the bit level assignment of the data bits as shown in Fig. 4, we apply the inverse of  $\Pi$  to the data bits before encoding them, specifically, we apply  $\Pi_{\text{data}}^{-1}$ , which is given by

$$V_0^{21600} = B_{3,0}^{\frac{n}{3}} \quad (31)$$

$$V_{21600}^{43200} = B_{2,0}^{\frac{n}{2}} \quad (32)$$

$$V_{43200}^{48600} = D_{\frac{k}{9}}^{\frac{k}{9}}. \quad (33)$$

Since the PR interleaver copies the parity bits  $C^{\frac{n}{4}}$  to bit level 1, the chain  $\Pi_{\text{data}}^{-1}$ , rate 3/4 DVB-S2 encoder and  $\Pi$  realizes the bit level assignment in Fig. 4.

### E. Decoder

We consider one channel use. The bit label  $\mathbf{B}$  gets mapped to the signal point  $X = x_{\mathbf{B}}$  at the input of the AWGN channel (5). The output  $Y = X + Z$  is quantized by a 5 bit quantizer, and the soft-demapper uses the quantizer output  $Q$  to calculate the soft information for each bit level  $i = 1, 2, 3$ . The soft information is

$$I_i = L_i + \pi_i \quad (34)$$

where  $L_i$  and  $\pi_i$  are the log-likelihood ratio and the a-priori information, respectively. The log-likelihood ratios are calculated as

$$L_i = \ln \frac{P_{Q|B_i}(Q|0)}{P_{Q|B_i}(Q|1)} \quad (35)$$

where

$$P_{Q|B_i}(Q|a) = \sum_{\mathbf{b} \in \{0,1\}^3: b_i=a} \frac{P_{QB_1\mathbf{s}}(Q\mathbf{b})}{P_{B_i}(a)} \quad (36)$$

with the marginals  $P_{B_1}$  and

$$P_{B_2}(a) = \sum_{b \in \{0,1\}} P_{\mathcal{S}}(ab) \quad (37)$$

$$P_{B_3}(a) = \sum_{b \in \{0,1\}} P_{\mathcal{S}}(ba). \quad (38)$$

The a-priori information for bit level  $i$  is

$$\pi_i = \ln \frac{P_{B_i}(0)}{P_{B_i}(1)}. \quad (39)$$

### F. Numerical Results

For SS-BMD and BS-BICM, we use an arithmetic distribution matcher with controllable overflow [23]. For a detailed discussion of the arithmetic matcher, we refer the reader to the extended version of this paper, which is in preparation. The results for SS-BMD, BS-BICM and uniform BICM are displayed in Fig. 6. We observe that the SNR gains of SS-BMD and BS-BICM over uniform BICM fit well to the asymptotic gains promised by the power-rate curves in Fig. 1. This suggests that we could effectively remove the entire shaping gap by using SS-BMD.

## VI. CONCLUSIONS

In this work, we showed that the CM capacity of the AWGN channel can effectively be achieved by a bit-metric decoder without iterative demapping-decoding. Our simulation results show that by combining Gray labeling, a distribution matcher and a binary code with a systematic encoder, the shaping gap can be removed. The remaining gap to capacity is because of the imperfections of the employed code. We observed that the interleaver has a strong impact on the performance. This suggests that interleaver and code should be designed together taking the varying reliabilities of the bit levels into account. To achieve very low error probabilities, LDPC codes with low error floor should be combined with zero error distribution matchers. The design of such distribution matchers is part of our current research. The techniques developed in this work may be useful in other scenarios, such as peak power constraints, fading channels, multiple antennas, and multiple users.

## VII. ACKNOWLEDGMENT

The author would like to thank L. Barletta, M. Stinner and G. Kramer for support in this work. This work was supported by the German Ministry of Education and Research in the framework of an Alexander von Humboldt Professorship.

APPENDIX

We prove the theorem by random coding arguments. We use letter typicality as defined in [24, Chap. 1].  $\mathcal{T}_\epsilon^n(P_{BY})$  is the set of sequences  $\mathbf{b}^n, y^n$  that are jointly  $\epsilon$ -typical with respect to  $P_{BY}$ . The set of conditionally typical sequences is defined as

$$\mathcal{T}_\epsilon^n(P_{BY}|y^n) := \{\mathbf{b}^n : (\mathbf{b}^n, y^n) \in \mathcal{T}_\epsilon^n(P_{BY})\}. \quad (40)$$

*Codebook Construction:* Choose  $2^{nR}$  codewords of length  $n$  by choosing the  $n \cdot 2^{nR}$  symbols according to  $P_B$ . Denote the resulting codebook by  $\mathcal{C}$ .

*Encoding:* Given message  $w \in [1 : 2^{nR}]$ , transmit  $\mathbf{b}^n(w)$ .

*Decoding:* For  $\epsilon_1 > \epsilon \geq 0$ , we define the bit metric

$$q_i(b_i^n, y^n) = \begin{cases} 1, & b_i^n \in \mathcal{T}_{\epsilon_1}^n(P_{B_iY}|y^n) \\ 0, & \text{otherwise.} \end{cases} \quad (41)$$

The corresponding decoding metric is

$$q(\mathbf{b}^n, y^n) = \prod_{i=1}^m q_i(b_i^n, y^n). \quad (42)$$

We define the set  $\hat{\mathcal{B}}(y^n) := \{\mathbf{b}^n \in \mathcal{C} : q(\mathbf{b}^n, y^n) = 1\}$ . The decoder output is

$$\begin{cases} \mathbf{b}^n, & \text{if } \mathcal{B}(y^n) = \{\mathbf{b}^n\} \\ \text{error,} & \text{otherwise.} \end{cases} \quad (43)$$

*Analysis:* Suppose message  $w$  was encoded. The two error events are

$$\mathcal{E}_1 := \{\mathbf{B}^n(w) \notin \hat{\mathcal{B}}(Y^n)\} \quad (44)$$

$$\mathcal{E}_2 := \{\exists \tilde{w} \neq w : \mathbf{B}^n(\tilde{w}) \in \hat{\mathcal{B}}(Y^n)\}. \quad (45)$$

*First error event:* The random experiment for  $\mathcal{E}_1$  has the distribution  $P_{BY}^n$ . We have

$$\begin{aligned} \Pr(\mathcal{E}_1) &= 1 - \Pr[q(\mathbf{B}^n, Y^n) = 1] \\ &= 1 - \Pr\left[\bigcap_{i=1}^m \{B_i^n \in \mathcal{T}_{\epsilon_1}^n(P_{B_iY}|Y^n)\}\right] \\ &\stackrel{(a)}{\leq} 1 - \Pr[(\mathbf{B}^n, Y^n) \in \mathcal{T}_{\epsilon_1}^n(P_{BY})] \stackrel{(b)}{n \rightarrow \infty} 0 \end{aligned} \quad (46)$$

where (a) follows because joint typicality implies marginal typicality [24, Sec. 1.5]. The limit (b) follows by [24, Theo. 1.1].

*Second error event:* The random experiment for  $\mathcal{E}_2$  has the distribution  $P_B^n P_Y^n$ . The probabilities of  $Y^n \in \mathcal{T}_\epsilon^n(P_Y)$  and  $\mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)$  approach 1 as  $n \rightarrow \infty$ , by [24, Theo. 1.1]. It therefore suffices to analyze for  $y^n \in \mathcal{T}_\epsilon^n(P_Y)$  the probability

$$\Pr[\mathcal{E}_2|Y^n = y^n, \mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)]. \quad (47)$$

By [24, Theo. 1.2], we have

$$|\mathcal{T}_{\epsilon_1}^n(P_{B_iY}|y^n)| \leq 2^{n\mathbb{H}(B_i|Y)(1+\epsilon_1)}. \quad (48)$$

The size of  $\hat{\mathcal{B}}(y^n)$  is thus bounded as

$$|\hat{\mathcal{B}}(y^n)| \leq 2^{n\sum_{i=1}^m \mathbb{H}(B_i|Y)(1+\epsilon_1)}. \quad (49)$$

By [24, Eq. (1.10) & (1.12)], we have

$$\Pr[\mathbf{B}^n = \mathbf{b}^n | \mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)] \leq \frac{2^{-n\mathbb{H}(\mathbf{B})(1-\epsilon)}}{1 - \delta_\epsilon(n)} \quad (50)$$

where  $\delta_\epsilon(n) \xrightarrow{n \rightarrow \infty} 0$ . We assume  $n$  is large enough such that  $\delta_\epsilon(n) \leq 1/2$ . The bound (50) then becomes

$$\Pr[\mathbf{B}^n = \mathbf{b}^n | \mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)] \leq 2 \cdot 2^{-n\mathbb{H}(\mathbf{B})(1-\epsilon)} \quad (51)$$

We have

$$\begin{aligned} &\Pr[\mathcal{E}_2|Y^n = y^n, \mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)] \\ &\leq (2^{nR} - 1) \sum_{\mathbf{b}^n \in \hat{\mathcal{B}}(y^n)} \Pr[\mathbf{B}^n = \mathbf{b}^n | \mathbf{B}^n \in \mathcal{T}_\epsilon^n(P_B)] \\ &\stackrel{(a)}{\leq} 2^{nR} \sum_{\mathbf{b}^n \in \hat{\mathcal{B}}(y^n)} 2 \cdot 2^{-n\mathbb{H}(\mathbf{B})(1-\epsilon)} \\ &\stackrel{(b)}{\leq} 2^{nR} 2^{n\sum_{i=1}^m \mathbb{H}(B_i|Y)(1+\epsilon_1)} \cdot 2 \cdot 2^{-n\mathbb{H}(\mathbf{B})(1-\epsilon)} \end{aligned} \quad (52)$$

where (a) follows by (51) and where we used (49) in (b). The term in (52) goes to zero for  $n \rightarrow \infty$  if

$$R + \left[ \sum_{i=1}^m \mathbb{H}(B_i|Y)(1 + \epsilon_1) \right] - \mathbb{H}(\mathbf{B})(1 - \epsilon) < 0. \quad (53)$$

Using  $\sum_{i=1}^m \mathbb{H}(B_i|Y) \leq m$  and  $\mathbb{H}(\mathbf{B}) \leq m$ , we have

$$\begin{aligned} R &< \sum_{i=1}^m [\mathbb{H}(B_i) - \mathbb{H}(B_i|Y)] \\ &\quad - \left\{ \left[ \sum_{i=1}^m \mathbb{H}(B_i) \right] - \mathbb{H}(\mathbf{B}) \right\} - m(\epsilon_1 + \epsilon) \\ &\Leftrightarrow R < \sum_{i=1}^m \mathbb{I}(B_i; Y) \\ &\quad - \left\{ \left[ \sum_{i=1}^m \mathbb{H}(B_i) \right] - \mathbb{H}(\mathbf{B}) \right\} - m(\epsilon_1 + \epsilon) \end{aligned} \quad (54)$$

for any  $0 < \epsilon < \epsilon_1$ . Thus, for any rate  $R$  less than (2), it follows by (46) and (54) that the probability of decoding error can be made as small as desired by choosing  $n$  large enough.

REFERENCES

- [1] E. Zehavi, "8-PSK trellis codes for a Rayleigh channel," *IEEE Trans. Commun.*, vol. 40, no. 5, pp. 873–884, 1992.
- [2] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, 2009.
- [3] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, 1997.
- [4] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 927–946, 1998.
- [5] A. G. i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2010, pp. 1–5.
- [6] G. Böcherer, F. Altenbach, A. Alvarado, S. Corroy, and R. Mathar, "An efficient algorithm to calculate BICM capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2012.
- [7] European Telecommun. Standards Inst. (ETSI) EN 302 307 v1.2.1, *Digital Video Broadcasting 2nd Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications*, 2009. [Online]. Available: [www.dvb.org](http://www.dvb.org)

- [8] M. C. Valenti and X. Xiang, "Constellation shaping for bit-interleaved LDPC coded APSK," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2960–2970, 2012.
- [9] J. Hossain, A. Alvarado, and L. Szczecinski, "BICM transmission using non-uniform qam constellations: Performance analysis and design," in *Proc. IEEE Int. Conf. Commun. (ICC)*. IEEE, 2010, pp. 1–6.
- [10] S. Y. Le Goff, B. K. Khoo, C. C. Tsimenidis, and B. S. Sharif, "Constellation shaping for bandwidth-efficient turbo-coded modulation with iterative receiver," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2223–2233, 2007.
- [11] R. Fischer, J. Huber, and U. Wachsmann, "On the combination of multilevel coding and signal shaping," in *Proc. Int. ITG Conf. Source Channel Coding*. Citeseer, 1998.
- [12] E. Agrell and A. Alvarado, "Optimal alphabets and binary labelings for BICM at low SNR," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6650–6672, 2011.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., 1968.
- [14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," <http://cvxr.com/cvx>, Sep. 2013.
- [15] G. Böcherer, "BACM—an algorithm for calculating BICM capacity." [Online]. Available: <http://www.georg-boecherer.de/bacm>
- [16] —, "Capacity-achieving probabilistic shaping for noisy and noiseless channels," Ph.D. dissertation, RWTH Aachen University, 2012. [Online]. Available: <http://www.georg-boecherer.de/capacityAchievingShaping.pdf>
- [17] G. Böcherer and R. Mathar, "Operating LDPC codes with zero shaping gap," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2011.
- [18] G. Ungerböck, "Channel coding with multilevel/phase signals," *Information Theory, IEEE Transactions on*, vol. 28, no. 1, pp. 55–67, 1982.
- [19] L. Ozarow and A. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1426–1428, 1990.
- [20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.
- [21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [22] Y. Li and W. E. Ryan, "Bit-reliability mapping in LDPC-coded modulation systems," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 1–3, 2005.
- [23] S. Baur and G. Böcherer, "Arithmetic distribution matching," Jan. 2014. [Online]. Available: <http://www.georg-boecherer.de/adm>
- [24] G. Kramer, *Topics in Multi-User Information Theory*. now publishers inc, 2008.