Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation

Georg Böcherer, *Member, IEEE*, Fabian Steiner, *Student Member, IEEE*, and Patrick Schulte

Abstract—A new coded modulation scheme is proposed. At the transmitter, the concatenation of a distribution matcher and a systematic binary encoder performs probabilistic signal shaping and channel coding. At the receiver, the output of a bitwise demapper is fed to a binary decoder. No iterative demapping is performed. Rate adaption is achieved by adjusting the input distribution and the transmission power. The scheme is applied to bipolar amplitude-shift keying (ASK) constellations with equidistant signal points and it is directly applicable to two-dimensional quadrature amplitude modulation (QAM). The scheme is implemented by using the DVB-S2 low-density parity-check (LDPC) codes. At a frame error rate of 10^{-3}, the new scheme operates within less than 1.1 dB of the AWGN capacity $\frac{1}{2}\log_2(1+\text{SNR})$ at any spectral efficiency between 1 and 5 bits/s/Hz by using only 5 modes, i.e., 4-ASK with code rate 2/3, 8-ASK with 3/4, 16-ASK and 32-ASK with 5/6, and 64-ASK with 9/10.

Index Terms—Rate adaptation, Probabilistic shaping, Bandlimited communication, Modulation, Amplitude-shift keying, Quadrature amplitude modulation, Forward error correction.

I. INTRODUCTION

RELIABLE communication over the *additive white Gaussian noise* (AWGN) channel is possible if the transmission rate per real dimension does not exceed the capacity-power function

$$\mathsf{C}(\mathsf{P}) = \frac{1}{2}\log_2(1+\mathsf{P}/1) \qquad (1)$$

where P is the transmission power and $\mathsf{P}/1$ is the *signal-to-noise ratio* (SNR). Throughout this work, we assume unit variance noise and we change the SNR by varying the transmission power P. An important consequence of (1) is that to achieve power and bandwidth efficient transmission over a substantial range of SNRs, a communication system must adapt its transmission rate to the SNR. Contemporary standards such
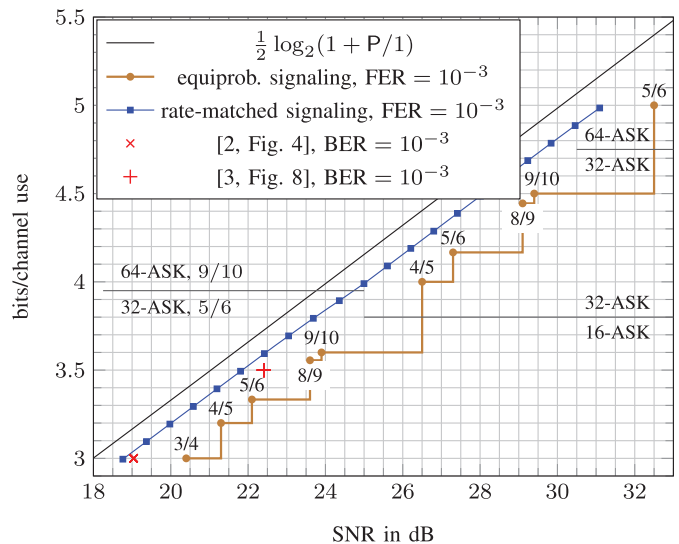
Fig. 1. Comparison of rate-matching by non-equiprobable signaling to conventional equiprobable signaling. Both schemes use LDPC codes from the DVB-S2 standard. Gains of up to 4 dB (at a spectral efficiency of 4.6 bits/s/Hz) are obtained. The rate-matched scheme uses two modes, namely 64-ASK with a rate 9/10 code and 32-ASK with a rate 5/6 code. Equiprobable signaling uses 10 modes, which combine {16, 32, 64}-ASK with {3/4, 4/5, 5/6, 8/9, 9/10} rate codes. The two schemes are evaluated at a *frame error rate* (FER) of 10^{-3}. For comparison, the operating points of the turbo coded modulation schemes with probabilistic shaping suggested in [2] (red ×) and [3] (red +) are displayed. The operating points of [2], [3] have a *bit error rate* (BER) of 10^{-3}. Since BER ≤ FER, our rate-matched scheme is more energy efficient than the schemes suggested in [2], [3], see Sec. II-A for a further discussion.

as DVB-S2 [1] support input constellations of different sizes and *forward error correction* (FEC) at various code rates. The combination of a constellation size with a code rate forms a transmission mode. By choosing the appropriate mode, the system adapts to the SNR. In Fig. 1 we display the operating points where a *frame error rate* (FER) less or equal to 10^{-3} can be achieved. The DVB-S2 *low-density parity-check* (LDPC) codes are used with *bipolar amplitude-shift keying* (ASK) constellations, i.e., the signal points are equidistant and symmetric around the origin[1]. Observe that the SNR gap between the operating points and the capacity-power function varies with the SNR. At the upper corner points, the gap is between 2 and 2.5 dB. Two factors contribute to this gap: first, the LDPC codes

---

[1]In the DVB-S2 standard, complex *amplitude phase shift keying* (APSK) constellations are used because of the non-linear satellite channel [4]. In this work, we consider the real linear AWGN channel, for which bipolar ASK constellations work well.

have finite length (coding gap) and second, the input distribution is uniform (shaping gap). At the lower corner points, the gap can be as large as 4.5 dB because the system resorts to one of the supported transmission modes.

To provide modes with a finer granularity, we can increase the number of code rates and constellation sizes. This approach is taken in the extension DVB-S2X [5] of DVB-S2. The system complexity increases with the number of supported code rates, which suggests the use of rate-compatible codes [6]. In [7]–[10], rate-compatible LDPC codes are designed.

In this work, we take a different approach and propose a new coded modulation scheme that uses probabilistic shaping to solve the problem of coarse mode granularity and to remove the shaping gap. With ASK constellations and only one LDPC code rate per constellation size, our scheme operates within less than 1.1 dB of capacity over the whole considered range of SNR, see Fig. 1. For the transmitter, we propose a new scheme, which we call *probabilistic amplitude shaping* (PAS). The PAS scheme concatenates a distribution matcher [11] for probabilistic shaping with a systematic binary encoder for FEC. At the receiver, bit-metric decoding [12]–[14] is used. No iterative demapping is required. Our scheme directly applies to two-dimensional *quadrature amplitude modulation* (QAM) constellations by mapping two real ASK symbols to one complex QAM symbol. We presented a preliminary version of PAS in [15].

This work is organized as follows. In Sec. II, we give an overview of the related literature on coded modulation with probabilistic shaping. In Sec. III, we review the information theoretic limits of the AWGN channel and we discuss optimal signaling for ASK constellations. We introduce PAS in Sec. IV and we show in Sec. V how PAS can be implemented by using distribution matching. In Sec. VI, we combine PAS at the transmitter with bit-metric decoding at the receiver. We discuss in Sec. VII the code design problem posed by our scheme and present a bit-mapper optimization heuristic. We propose a rate adaption scheme in Sec. VIII and we discuss numerical results in Sec. IX. The work is concluded in Sec. X.

II. Related Literature

On the AWGN channel, uniformly distributed inputs can be up to 1.53 dB less power efficient than Gaussian inputs [16, Sec. IV.B], see also Sec. III-D of this work. Coded modulation with probabilistic shaping uses a non-uniform distribution on equidistant signal points to overcome the shaping gap. For works with a theoretical focus on the topic, see for example [17], [18] and references therein. Here, we review some of the existing works that address the practical implementation of probabilistic shaping.

A. Gallager's Scheme

Gallager proposes in [19, p. 208] to use a many-to-one mapping to turn a channel with non-uniform inputs into a super channel with uniform inputs. This approach is combined with turbo coding in [2] and [3]. In [20], Gallager's scheme is combined with convolutional codes. Optimal mappings are investigated in [21], see also [22, Sec. I.B]. When Gallager's

scheme is combined with binary FEC, undoing the many-to-one mapping at the receiver is challenging. In [2], [3], [20], this inversion is achieved by iterative demapping, which increases the system complexity compared to uniform signaling. In [3], the authors choose the many-to-one mapping according to the desired transmission rate. The resulting rate granularity is coarse. To achieve a finer granularity, more bits must be mapped to the same signal point, which increases the system complexity further.

B. Trellis Shaping

Trellis shaping is proposed in [23] and it is discussed in detail in [24, Chap. 4] and [25, Sec. 4.4]. The transmitter first selects a set of sequences. A shaping code then selects from this set the sequence of minimum energy for transmission. The authors of [26] combine multilevel coding with trellis shaping using convolutional codes, see also [27, Sec. VIII]. *Bit-interleaved coded modulation* (BICM) [28], [29] is used in [30] for lower bit levels and a convolutional shaping code selects the two highest bit-levels for energy minimization. LDPC codes were used as shaping codes in [31]. Since shaping is achieved by the decoder of a shaping code, it is difficult to make trellis shaping flexible to support different shaping rates.

C. Shell Mapping

Shell mapping [32] was proposed independently by several groups around the same time, see [33, Sec. VII.], [24, Chap. 8]. In shell mapping, the transmitter first selects a shell, which is a set of low energy input sequences. It then selects a sequence within the selected shell for transmission. Shell mapping combined with trellis coding is used in the ITU-T Recommendation V.34 [34]. The main challenge is to efficiently index the shells.

D. Superposition Coding

Superposition coding was proposed in [35]. The transmitter uses several encoders. Each encoder maps its output to a signal point. The selected signal points are then superposed and transmitted. The receiver uses multistage decoding. Turbo codes are considered in [35] and variations of superposition coding are proposed and analyzed in [36]. The author of [37] uses LDPC codes and proposes signal constellations that reduce the number of decoding stages. Superposition coding requires sequential decoding of the individual codes or iterative demapping.

E. Concatenated Shaping

The authors of [38] concatenate BICM encoding with shaping. A turbo encoder is followed by a one-to-one shaping block code. The scheme is improved in [39] by iterative demapping. A variation of this scheme is presented in [40], where the one-to-one shaping block code is followed by a many-to-one mapping reminiscent of Gallager's scheme. The shaping scheme [39] is used in [41] with LDPC codes on *amplitude and phase-shift keying* (APSK) constellations. Since the shaping decoder is placed before the FEC decoder, only limited gains are observed in [38]. The works [39]–[41] therefore use iterative demapping.

F. Bootstrap Scheme

In [42], [43, Chap. 7], we proposed the bootstrap scheme, which separates shaping and FEC. First, the data is shaped, i.e., uniformly distributed data symbols are transformed into non-uniformly distributed symbols. The shaped data is then systematically encoded. The generated check bits are also shaped and embedded in the shaped data of the next transmission block. The bootstrap scheme is implemented in [42], [43, Chap. 7] using Geometric Huffman Coding [44] for shaping and LDPC codes from the DVB-S2 standard for FEC. Since shaping is done prior to FEC encoding, the bootstrap scheme borrows from the reverse concatenation proposed in [45], [46] for magnetic recording. The authors in [18] call the bootstrap scheme *chaining construction* and prove that it is capacity-achieving for any discrete memoryless channel. The drawback of the bootstrap scheme is that shaping is done over several consecutive transmission blocks, which have to be decoded in reverse (bootstrapped). This increases the latency and is prone to error propagation over blocks.

III. OPTIMAL SIGNALING OVER THE AWGN CHANNEL

The discrete time AWGN channel at time instance i is described by the input-output relation

$$Y_i = X_i + Z_i \qquad (2)$$

where the noise terms Z_i, $i = 1, 2, 3, \ldots$ are *independent and identically distributed* (iid) according to a zero mean Gaussian distribution with unit variance. For n_c channel uses, the input is subject to the power constraint

$$\frac{\mathbb{E}\left[\sum_{i=1}^{n_c} X_i^2\right]}{n_c} \leq \mathsf{P} \qquad (3)$$

where $\mathbb{E}[\cdot]$ denotes expectation.

A. AWGN Capacity

Let $\hat{X}^{n_c} = \hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{n_c}$ be the output of the decoder that estimates the input X^{n_c} from the output Y^{n_c}. The block error probability is $P_e = \Pr(\hat{X}^{n_c} \neq X^{n_c})$. The channel coding theorem [19, Theorem 7.4.2] states that by choosing n_c large enough, P_e can be made as small as desired if the transmission rate R is smaller than the channel capacity $\mathsf{C}(\mathsf{P})$. The goal of this work is to design a modulation scheme that achieves reliable transmission close to the capacity-power function (1), i.e., for **any** average signal power P, we want to reliably transmit data over the AWGN channel at a rate that is close to $\mathsf{C}(\mathsf{P})$.

B. Amplitude-Shift Keying

Let X be distributed on some finite alphabet $\mathfrak{X}$ of signal points and suppose $\mathbb{E}[X^2] = \mathsf{P}$. The channel coding theorem states that reliable transmission at rate R with average power P is possible if

$$R < \mathbb{I}(X; Y) \qquad (4)$$

where $\mathbb{I}(\cdot; \cdot)$ denotes the mutual information in bits. The first step in designing a coded modulation system is thus to choose an alphabet $\mathfrak{X}$ and a distribution P_X on $\mathfrak{X}$ such that $\mathbb{I}(X; Y)$ is close to the maximum value $\mathsf{C}(\mathsf{P})$. As input alphabet, we use an ASK constellation with 2^m signal points, namely

$$\mathfrak{X} = \{\pm 1, \pm 3, \ldots, \pm(2^m - 1)\}. \qquad (5)$$

We scale X by the *constellation scaling* $\Delta > 0$ and the resulting input/output relation is

$$Y = \Delta X + Z. \qquad (6)$$

The power constraint for X is now

$$\mathbb{E}[(\Delta X)^2] \leq \mathsf{P}. \qquad (7)$$

C. Optimization of the ASK Input

We use signal point $x \in \mathfrak{X}$ with probability

$$P_{X_\nu}(x) = A_\nu e^{-\nu x^2}, \qquad A_\nu = \frac{1}{\sum_{x' \in \mathfrak{X}} e^{-\nu x'^2}}. \qquad (8)$$

The scalar A_ν ensures that the probabilities assigned by P_{X_ν} add to 1. Distributions of the form (8) are called *Maxwell-Boltzmann distributions*, see, e.g., [47]. For a fixed constellation scaling Δ and power constraint P, we choose the input distribution

$$P_{X_\Delta}(x) = P_{X_\nu}(x) \text{ with } \nu \colon \mathbb{E}[(\Delta X_\nu)^2] = \mathsf{P}. \qquad (9)$$

The distribution P_{X_Δ} maximizes entropy subject to the power constraint (7), see [48, Chapter 12]. Some basic manipulations show that $\mathbb{E}[X_\nu^2]$ is strictly monotonically decreasing in ν. Thus, the value of ν for which the condition (9) is fulfilled can be found efficiently by using the *bisection method*. For each constellation scaling Δ, the distribution P_{X_Δ} satisfies the power constraint. We now maximize the mutual information over all input distributions from this family, i.e., we solve

$$\max_\Delta \mathbb{I}(X_\Delta; \Delta X_\Delta + Z). \qquad (10)$$

The mutual information $\mathbb{I}(X_\Delta; \Delta X_\Delta + Z)$ is a unimodal function of Δ and the optimization problem can be solved efficiently by using the *golden section method*. We denote the maximizing scaling by $\Delta^{\clubsuit}$ and the corresponding distribution by $P_{X^{\clubsuit}}$.

D. Shaping Gap

In Table I, we display the shaping gains of our input $X^{\clubsuit}$ as compared to a uniformly distributed input. For increasing rates and constellation sizes, the shaping gain increases and approaches the upper bound of $10 \log_{10} \frac{\pi e}{6} \approx 1.53$ dB [16, Sec. IV.B]. The bound can also be derived by probabilistic arguments, see [49, Comment 4]). For 64-ASK and a rate of 5 bits per channel use, the shaping gain is 1.33 dB.

Remark 1: The distribution $P_{X^{\clubsuit}}$ is suboptimal in general. The optimal distribution P_{X^*} can be approximated numerically by using the Blahut-Arimoto algorithm [50], [51]. However, there is no analytical expression of P_{X^*}. For the operating points in Table I, the energy efficiency of $X^{\clubsuit}$ is within 0.11 dB of capacity $\mathsf{C}(\mathsf{P})$, so the gain from optimizing further is bounded by 0.11 dB.

TABLE I
SNR GAPS OF UNIFORM ASK AND $X^{\clubsuit}$ TO CAPACITY C(P)

constellation	rate [$\frac{\text{bits}}{\text{channel use}}$]	capacity C(P) SNR [dB]	uniform ASK SNR [dB]	gap [dB]	$X^{\clubsuit}$ SNR [dB]	gap [dB]	shaping gain [dB]
4-ASK	1	4.7712	5.1181	0.3469	4.8180	0.0468	0.3001
8-ASK	2	11.7609	12.6187	0.8578	11.8425	0.0816	0.7762
16-ASK	3	17.9934	19.1681	1.1747	18.0910	0.0976	1.0771
32-ASK	4	24.0654	25.4140	1.3486	24.1706	0.1052	1.2434
64-ASK	5	30.0988	31.5384	1.4396	30.2078	0.1090	1.3306

E. Input Distribution: Operational Meaning

Suppose we use the channel n_c times, i.e., we transmit length n_c codewords. We use codeword x^{n_c} with probability $P_{X^{n_c}}(x^{n_c})$. If our transmission rate R is larger than the average mutual information between input and output, i.e., if

$$R > \frac{\sum_{i=1}^{n_c} \mathbb{I}(X_i; Y_i)}{n_c} \tag{11}$$

then the probability of error $\Pr(\hat{X}^n \neq X^n)$ is bounded away from zero for any decoder. This result follows as a corollary of the converse of the channel coding theorem [19, Theorem 7.3.1].

The mutual information terms $\mathbb{I}(X_i; Y_i)$ are calculated according to the marginal distributions P_{X_i}. We call the condition (11) the *Coded Modulation Converse*. The insight we take from (11) is that building a transmitter with marginal input distributions

$$P_{X_i} \approx P_{X^{\clubsuit}} \tag{12}$$

is necessary to achieve reliable transmission close to the rates displayed for $X^{\clubsuit}$ in Table I.

IV. PROBABILISTIC AMPLITUDE SHAPING

We now develop a transmitter with property (12).

A. Preliminaries

We make the following two observations:

1) Amplitude-Sign Factorization: We can write $X^{\clubsuit}$ as

$$X^{\clubsuit} = A \cdot S \tag{13}$$

where $A = |X^{\clubsuit}|$ is the *amplitude* of the input and where $S = \text{sign}(X^{\clubsuit})$ is the *sign* of the input. By (5), the amplitudes take values in

$$\mathcal{A} := \{1, 3, \ldots, 2^m - 1\}. \tag{14}$$

We see from (8) that the distribution $P_{X^{\clubsuit}}$ is symmetric around zero, i.e., we have

$$P_{X^{\clubsuit}}(x) = P_{X^{\clubsuit}}(-x) \tag{15}$$

and therefore, A and S are stochastically independent and S is uniformly distributed, i.e., we have

$$P_{X^{\clubsuit}}(x) = P_A(|x|) \cdot P_S(\text{sign}(x)), \quad \forall x \in \mathcal{X} \tag{16}$$

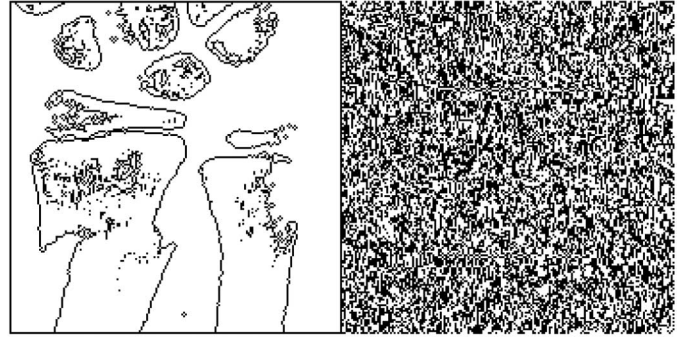$$P_S(1) = P_S(-1) = \frac{1}{2}. \tag{17}$$

Fig. 2. The black and white pixels of a 180×180 picture are represented by 1s and 0s, respectively, and then encoded by a DVB-S2 rate 1/2 LDPC code. The resulting check bits are displayed next to the picture. The empirical distribution of the original picture is $P_{\bar{D}}(1) = 1 - P_{\bar{D}}(0) = 0.1082$ and the empirical distribution of the check bits is $P_{\bar{R}}(1) = 1 - P_{\bar{R}}(0) = 0.4970$.

2) Uniform Check Bit Assumption: The second observation is on systematic binary encoding. A systematic generator matrix of an (n, k) binary code has the form

$$G = [I_k | P] \tag{18}$$

where I_k is the $k \times k$ identity matrix and P is a $k \times (n - k)$ matrix. P is the *parity matrix* of the code. The generator matrix G maps k data bits D^k to a length n codeword via

$$D^k G = (D^k | R^{n-k}) \tag{19}$$

where R^{n-k} are redundant bits that are modulo-two sums of data bits. Suppose the data bits have some distribution P_{D^k}. Since the encoding is systematic, this distribution is preserved at the output of the encoder. What is the distribution of the redundancy bits? To address this question, consider two independent data bits D_1 and D_2. The modulo-two sum $R = D_1 \oplus D_2$ is then more uniformly distributed than the individual summands D_1 and D_2, see the appendix for the proof. This suggests that if the redundancy bits are the modulo-two sum of a large enough number of data bits, then their distribution is close to uniform. An example of this phenomenon is shown in Fig. 2. We therefore follow [52, Sec. VI], [53, Chap. 5], [43, Sec. 7.1] and assume in the following that the redundancy bits are uniformly distributed and we call this assumption the *uniform check bit assumption*.

B. Encoding Procedure

Consider block transmission with n_c symbols from a 2^m-ASK constellation. Since we use binary error correcting codes, we
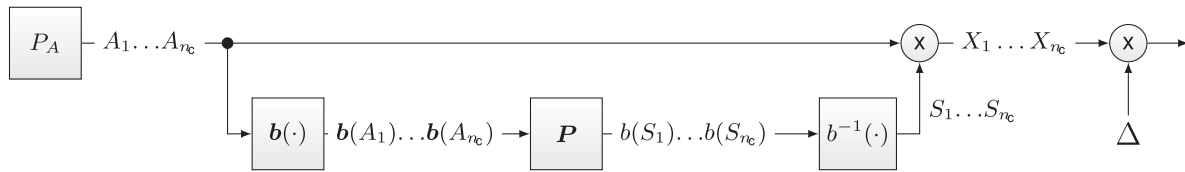
Fig. 3. PAS. The ASK amplitudes A_i take values in $\mathcal{A} = \{1, 3, \ldots, 2^m - 1\}$. The amplitudes A_i are represented by their binary labels $\boldsymbol{b}(A_i)$. Redundancy bits $b(S_i)$ result from multiplying the binary string $\boldsymbol{b}(A_1)\boldsymbol{b}(A_2) \cdots \boldsymbol{b}(A_{n_c})$ by the parity matrix $\boldsymbol{P}$ of a systematic generator matrix $[\boldsymbol{I}_k | \boldsymbol{P}]$. The redundancy bits $b(S_i)$ are transformed into signs S_i and multiplied with the amplitudes A_i. The resulting signal points $X_i = A_i S_i$ take values in $\mathcal{X} = \{\pm 1, \pm 3, \ldots, \pm(2^m - 1)\}$. The signal points X_i are scaled by Δ and ΔX_i is transmitted over the channel.

label each of the 2^{m-1} amplitudes by a binary string of length $m - 1$ and we label each of the signs ± 1 by a bit, i.e., we use

$$A \mapsto \boldsymbol{b}(A) \in \{0, 1\}^{m-1} \qquad (20)$$

$$S \mapsto b(S) \in \{0, 1\}. \qquad (21)$$

For the sign, we use $b(-1) = 0$ and $b(1) = 1$. We discuss the choice of $\boldsymbol{b}(A)$ in Sec. VI-C. We use a rate $k/n = (m-1)/m$ binary code with systematic generator matrix $\boldsymbol{G} = [\boldsymbol{I}_k | \boldsymbol{P}]$. For block transmission with n_c channel uses, the block length of the code is $n = n_c m$ and the dimension of the code is $k = n_c(m - 1)$. The encoding procedure is displayed in Fig. 3. It works as follows.

1) A *discrete memoryless source* (DMS) $\boxed{P_A}$ outputs amplitudes $A_1, A_2, \ldots, A_{n_c}$ that are iid according to P_A. We explain in Sec. V how the DMS $\boxed{P_A}$ can be emulated from binary data by distribution matching.
2) Each amplitude A_i is represented by its label $\boldsymbol{b}(A_i)$.
3) The resulting length $(m-1)n_c = k$ binary string is multiplied by the parity matrix $\boldsymbol{P}$ to generate $n - k = n_c$ sign labels $b(S_1), b(S_2), \ldots, b(S_{n_c})$.
4) Each sign label $b(S_i)$ is transformed into the corresponding sign S_i.
5) The signal $X_i = A_i \cdot S_i$ is scaled by Δ and transmitted.

We call this procedure *probabilistic amplitude shaping* (PAS). Since the signs S^{n_c} are a deterministic function of the amplitudes A^{n_c}, the input symbols $X_1, X_2, \ldots, X_{n_c}$ are correlated. Under the uniform check bit assumption, the marginal distributions are

$$P_{X_i}(x_i) = P_A(|x_i|) P_{S_i}(\text{sign}(x_i)) \qquad (22)$$

$$= P_A(|x_i|) \frac{1}{2} \qquad (23)$$

$$= P_{X^\clubsuit}(x_i) \qquad (24)$$

that is, if the uniform check bit assumption holds, then PAS has the desired property (12).

Remark 2: PAS is a special case of the bootstrap scheme [18], [42]. After encoding, the redundancy bits are approximately uniformly distributed. Instead of transforming them into a sequence of symbols with a non-uniform distribution and transmitting them in the next block as is done in the bootstrap scheme, they are left unchanged, since by (17), the uniform distribution is desirable.
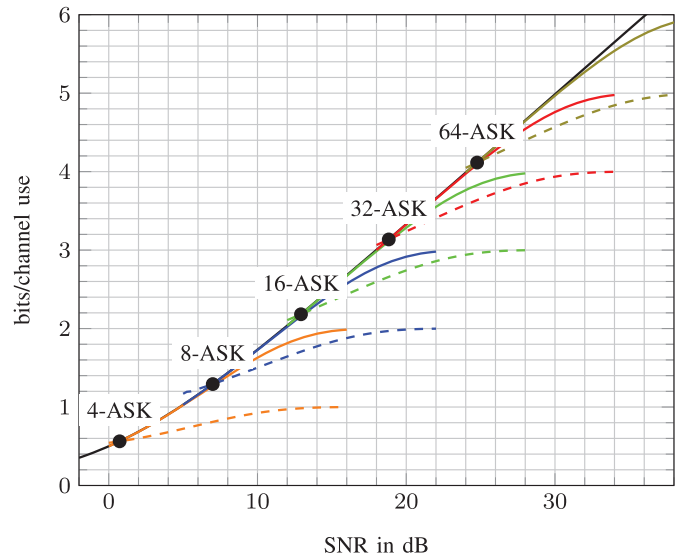

Fig. 4. The mutual information curves (solid) and the transmission rate curves (dashed) for ASK. The optimal operating points for rate $(m-1)/m$ codes are indicated by dots.

C. Optimal Operating Points

We study the rates at which reliable transmission is possible with our scheme. By (11), reliable communication at rate R is achievable only if

$$R < \frac{\sum_{i=1}^{n_c} \mathbb{I}(X_i; Y_i)}{n_c} = \mathbb{I}(X; Y) = \mathbb{I}(AS; Y). \qquad (25)$$

Since A^{n_c} represents our data, our transmission rate is

$$R = \frac{\mathbb{H}(A^{n_c})}{n_c} = \mathbb{H}(A) \quad \left[\frac{\text{bits}}{\text{channel use}}\right] \qquad (26)$$

and condition (25) becomes

$$\mathbb{H}(A) < \mathbb{I}(AS; Y). \qquad (27)$$

In Fig. 4, both the mutual information $\mathbb{I}(AS; Y)$ (solid lines) and transmission rate $\mathbb{H}(A)$ (dashed lines) are displayed for 4, 8, 16, 32, and 64-ASK. For high enough SNR, the mutual information saturates at m bits and the transmission rate saturates at $m - 1$ bits. Optimal error correction for block length $n_c \to \infty$ would operate where the transmission rate curve crosses the mutual information curve. These crossing points are indicated by dots in Fig. 4. How to operate at other transmission rates is the topic of rate adaption and we discuss this in detail in Sec. VIII.
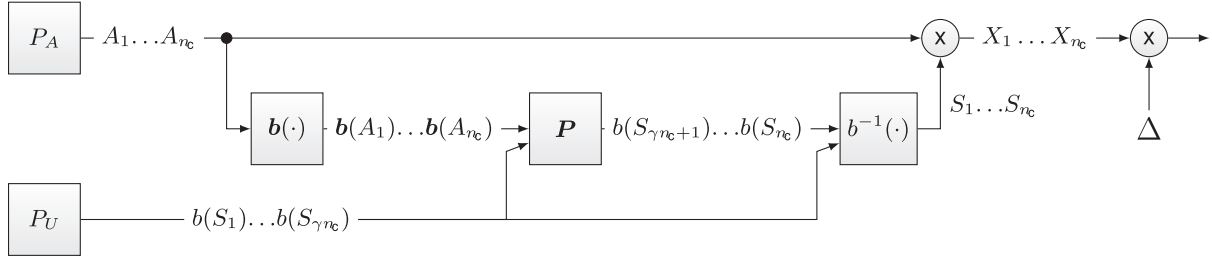
Fig. 5. Extension of PAS to code rates higher than $(m-1)/m$. The fraction γ of the signs is used for data, which is modelled as the output of a Bernoulli-1/2 DMS P_U.

D. PAS for Higher Code Rates

We observe in Fig. 4 that the ASK mutual information curves stay close to the capacity $C(P)$ over a certain range of rates above the optimal operating points. We therefore extend our PAS scheme to enable the use of code rates higher than $(m-1)/m$ on 2^m-ASK constellations. We achieve this by using some of the signs S_i for uniformly distributed data bits. We illustrate this extension of the PAS scheme in Fig. 5. Let γ denote the fraction of signs used for data bits. We interpret γn_c uniformly distributed data bits as sign labels $b(S_1) \cdots b(S_{\gamma n_c})$. These γn_c bits and the $(m-1)n_c$ bits from the amplitude labels are encoded by the parity matrix of a rate c code, which generates the remaining $(1-\gamma)n_c$ sign labels. The code rate can be expressed in terms of m and γ as

$$c = \frac{m-1+\gamma}{m}. \tag{28}$$

For a given code rate c, the fraction γ is given by

$$\gamma = 1 - (1-c)m. \tag{29}$$

Since a fraction γ of the signs now carries information, the transmission rate of the extended PAS scheme is given by

$$R = \frac{\mathbb{H}(A^{n_c}) + \mathbb{H}(S^{\gamma n_c})}{n_c} = \mathbb{H}(A) + \gamma \left[\frac{\text{bits}}{\text{channel use}}\right]. \tag{30}$$

The optimal operating point is then given by the crossing of the rate curve $\mathbb{H}(A) + \gamma$ and the mutual information. In Fig. 6, we display for 8-ASK the optimal operating points for $c = 2/3$ and $c = 3/4$.

V. DISTRIBUTION MATCHING

In a digital communication system, usually a binary interface separates the source coding part from the channel coding part [54, Chap. 1]. Up to now, our scheme does not have such a binary interface, since some of our data is output by an amplitude source $\boxed{P_A}$. We therefore add a device that takes as input k_c uniformly distributed independent bits U^{k_c} and outputs n_c amplitudes $\tilde{A}^{n_c}$. We illustrate this in Fig. 7. We require the following properties for our device:

P1 The input provides a binary interface to the source coding part of the digital communication system.

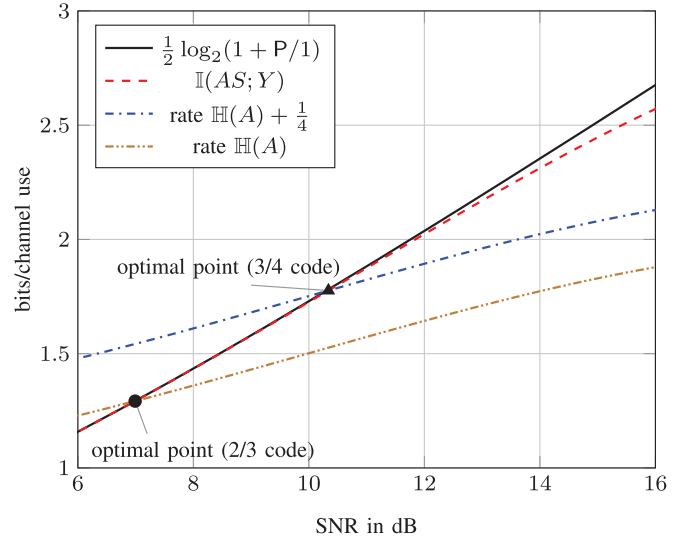P2 The input can be recovered from the output, i.e., the mapping is invertible.

Fig. 6. Optimal operating points of 8-ASK for PAS ($c = 2/3$) and extended PAS ($c = 3/4$).

Fig. 7. The matcher transforms uniform data blocks of length k_c into n_c amplitudes that are approximately distributed according to the desired distribution P_A. By replacing the amplitude source P_A in the PAS diagrams in Fig. 3 and Fig. 5 by a matcher, our scheme provides a binary interface to the source coding part of a digital communication system.

P3 The output is approximately the output of a DMS $\boxed{P_A}$.

A device with such properties is called a *distribution matcher* [44]. Variable length distribution matchers were proposed in [16, Sec. IV.A], [47, Sec. VII] and [55] and their design was studied in [43], [44], [47], [56]–[58].

Variable length matchers can lead to buffer overflow, synchronization loss and error propagation, see, e.g., [47, Sec. 1]. We therefore use the *constant composition distribution matcher* (CCDM) proposed in [11], which is a fixed length matcher.

A. Constant Composition Distribution Matching

1) Definitions: To describe the CCDM, we use *types*, which are described in detail, e.g., in [59], [60, Chap. 2], [48, Sec. 11.1], [61, Chap. 2]. Consider an amplitude sequence a^{n_c}. Let n_α be the number of times that amplitude α occurs in a^{n_c},

i.e., $n_\alpha = |\{i : a_i = \alpha\}|$. Let $P_{\bar{A}}$ be the empirical distribution of a^{n_c}, i.e.,

$$P_{\bar{A}}(\alpha) = \frac{n_\alpha}{n_c}, \quad \alpha \in \mathcal{A}. \tag{31}$$

Since each probability $P_{\bar{A}}(\alpha)$ is an integer multiple of $1/n_c$, the distribution $P_{\bar{A}}$ is called an n_c-*type*. The set $\mathcal{T}_{P_{\bar{A}}}^{n_c}$ contains all permutations of a^{n_c} and is called the *type class* of $P_{\bar{A}}$. A set of sequences is called a *constant composition code*, if all sequences in the set are of the same type.

Example 1: Consider the 4-ASK constellation $\mathcal{X} = \{\pm 1, \pm 3\}$ and the block length $n_c = 4$. The amplitudes are $\mathcal{A} = \{1, 3\}$. Suppose the desired distribution is

$$P_A(1) = \frac{3}{4}, \quad P_A(3) = \frac{1}{4}. \tag{32}$$

The probabilities are integer multiples of $1/4$, so P_A is a 4-type. The corresponding type class of length $n_c = 4$ sequences is

$$\mathcal{T}_{P_A}^4 = \{(1, 1, 1, 3), (1, 1, 3, 1), (1, 3, 1, 1), (3, 1, 1, 1)\} \tag{33}$$

where in each sequence, the amplitudes 1 and 3 occur $n_1 = 3$ and $n_3 = 1$ times, respectively. A CCDM maps length k_c binary strings to sequences in $\mathcal{T}_{P_A}^4$. There are 4 sequences in $\mathcal{T}_{P_A}^4$, so we have

$$k_c = \log_2 |\mathcal{T}_{P_A}^4| = 2. \tag{34}$$

The following lookup table defines a CCDM.

$$\begin{aligned}
00 &\mapsto (1, 1, 1, 3), \quad 01 \mapsto (1, 1, 3, 1), \\
10 &\mapsto (1, 3, 1, 1), \quad 11 \mapsto (3, 1, 1, 1).
\end{aligned} \tag{35}$$

This mapping maps binary strings to amplitude sequences and thereby provides a binary interface (Property P1). The mapping is one-to-one and therefore invertible (Property P2).

2) General Construction: Consider some block length n_c and a desired amplitude distribution P_A. We construct a CCDM to emulate the DMS $\boxed{P_A}$ as follows.

- We quantize P_A by an n_c-type distribution, i.e., we use a distribution $P_{\bar{A}}$ with

$$n_\alpha \approx n_c \cdot P_A(\alpha), \quad \alpha \in \mathcal{A}. \tag{36}$$

Quantizations with property (36) can be calculated easily, e.g., by rounding $n_c \cdot P_A(\alpha)$ to the nearest integer. We use the quantization rule [22, Algorithm 2] in our implementation. Note that in Example 1, the quantization step is not needed since the desired distribution is already of the required type.

- We calculate the input length k_c as

$$k_c = \lfloor \log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}| \rfloor = \left\lfloor \log_2 \frac{n_c!}{n_1! n_3! \cdots n_{2^m - 1}!} \right\rfloor \tag{37}$$

where $n_\alpha = P_{\bar{A}}(\alpha) \cdot n_c$, $\alpha \in \mathcal{A} = \{1, 3, \ldots, 2^m - 1\}$. The rounding $\lfloor \cdot \rfloor$ is necessary in general since the cardinality of the type class $\mathcal{T}_{P_{\bar{A}}}^{n_c}$ may not be a power of two. In Example 1, rounding is not necessary because the cardinality of the type class is a power of two.

- We pick a set $\mathcal{C}_{ccdm} \subseteq \mathcal{T}_{P_{\bar{A}}}^{n_c}$ with $|\mathcal{C}_{ccdm}| = 2^{k_c}$ sequences and define a bijective mapping f_{ccdm} from $\{0, 1\}^{k_c}$ to $\mathcal{C}_{ccdm}$. In Example 1, the cardinality of $\mathcal{T}_{P_{\bar{A}}}^{n_c}$ is 2^{k_c} so $\mathcal{C}_{ccdm} = \mathcal{T}_{P_{\bar{A}}}^{n_c}$.

Remark 3: PAS with CCDM can be seen as an instance of shell mapping. Since each amplitude sequence is of the same type, each signal x^{n_c} has the same power, i.e.,

$$\frac{\sum_{i=1}^{n_c} x_i^2}{n_c} = \frac{\sum_{i=1}^{n_c} a_i^2}{n_c} = \mathsf{P}. \tag{38}$$

That is, the signals are selected from a shell that contains the sequences of energy $n_c \mathsf{P}$. A similar signaling was used in [62] do derive finite length performance bounds for optimal codes for the AWGN channel.

B. Performance of CCDM

1) Rate: The rate of the CCDM is

$$\frac{k_c}{n_c} = \frac{\lfloor \log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}| \rfloor}{n_c} \quad \left[\frac{\text{bits}}{\text{amplitude}} \right]. \tag{39}$$

An upper bound on the rate is

$$\frac{\lfloor \log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}| \rfloor}{n_c} \leq \frac{\log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}|}{n_c} \overset{(a)}{\leq} \mathbb{H}(\bar{A}) \tag{40}$$

where (a) follows by [48, Theorem 11.1.3]. By [11, Eq. (22)], the rate is lower bounded by

$$\begin{aligned}
&\frac{\lfloor \log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}| \rfloor}{n_c} \\
&\geq \mathbb{H}(\bar{A}) - \frac{(|\mathcal{A}| - 1) \log_2(n_c + |\mathcal{A}| - 1)}{n_c} - \frac{1}{n_c}. \tag{41}
\end{aligned}$$

The bounds (40) and (41) can be evaluated numerically and are useful for system design. In Example 1, the rate is $1/2$, while the rate of the DMS that we want to emulate is $\mathbb{H}(A) = 0.8113$. By choosing the block length n_c large enough, the upper and lower bounds (40) and (41) can be made arbitrarily close to $\mathbb{H}(A)$, e.g., for $n_c = 10\,000$ and P_A of Example 1, the bounds evaluate to

$$\mathbb{H}(A) \geq \frac{k_c}{n_c} \geq \mathbb{H}(A) - 0.0014288. \tag{42}$$

The actual rate of CCDM for $n_c = 10\,000$ and P_A of Example 1 is

$$\frac{k_c}{n_c} = \frac{\lfloor \log_2 |\mathcal{T}_{P_{\bar{A}}}^{n_c}| \rfloor}{10\,000} = \frac{8106}{10\,000} = 0.8106 \tag{43}$$

which is within 7×10^{-4} bits of the desired rate $\mathbb{H}(A) = 0.8113$.

2) Informational Divergence: Let $\tilde{A}^{n_c} = f_{ccdm}(U^{k_c})$ with distribution $P_{\tilde{A}^{n_c}}$ on $\mathcal{C}_{ccdm}$ be the output of the CCDM and let $P_A^{n_c}$ be the output distribution of the DMS $\boxed{P_A}$ that we want to emulate. By [63, Theorem 1.2], a good measure for similarity

in the context of channel coding is the normalized informational divergence (also known as *Kullback-Leibler divergence* or *relative entropy* [48, Sec. 2.3]) of $P_{\tilde{A}^{n_c}}$ and $P_A^{n_c}$, which is given by

$$\frac{\mathbb{D}(P_{\tilde{A}^{n_c}} \| P_A^{n_c})}{n_c} = \frac{\displaystyle\sum_{a^{n_c} \in \mathcal{C}_{ccdm}} P_{\tilde{A}^{n_c}}(a^{n_c}) \log_2 \frac{P_{\tilde{A}^{n_c}}(a^{n_c})}{P_A^{n_c}(a^{n_c})}}{n_c}. \quad (44)$$

Informational divergence is non-negative and equal to zero if and only if the two compared distributions are equal [48, Theorem 2.6.3]. By [11, Eq. (17)], we can write (44) as

$$\frac{\mathbb{D}(P_{\tilde{A}^{n_c}} \| P_A^{n_c})}{n_c} = \mathbb{H}(\bar{A}) - \frac{k_c}{n_c} + \mathbb{D}(P_{\bar{A}} \| P_A). \quad (45)$$

In Example 1, we have $P_{\bar{A}} = P_A$, so $\mathbb{H}(\bar{A}) = \mathbb{H}(A)$ and $\mathbb{D}(P_{\bar{A}} \| P_A) = 0$ and the normalized informational divergence evaluates to

$$\frac{\mathbb{D}(P_{\tilde{A}^{n_c}} \| P_A^{n_c})}{n_c} = 0.8113 - \frac{1}{2} = 0.3113 \left[\frac{\text{bits}}{\text{amplitude}}\right]. \quad (46)$$

For $n_c = 10\,000$ and P_A of Example 1, we have

$$\frac{\mathbb{D}(P_{\tilde{A}^{n_c}} \| P_A^{n_c})}{n_c} = 0.8113 - 0.8106$$

$$= 7 \times 10^{-4} \quad \left[\frac{\text{bits}}{\text{amplitude}}\right]. \quad (47)$$

In [11], it is shown that the CCDM has Property P3 in the following sense:

1) As $n_c \to \infty$, the rate approaches $\mathbb{H}(A)$, i.e., we have

$$\frac{k_c}{n_c} \to \mathbb{H}(A). \quad (48)$$

2) As $n_c \to \infty$, the normalized informational divergence (44) approaches zero.

C. Efficient Implementation of CCDM

As we can see from our discussion of Example 1, to emulate a DMS $\boxed{P_A}$ well by CCDM, the block length n_c has to be large. The size of the code book $\mathcal{C}_{ccdm}$ grows exponentially with the input length k_c and the implementation of the CCDM by a lookup table as in (35) becomes infeasible because of memory limitations. In [11], we propose an algorithm based on arithmetic coding that performs the mapping f_{ccdm} without storing $\mathcal{C}_{ccdm}$. At [64], we provide a flexible CCDM implementation written in C/Matlab that takes n_c, P_A, and u^{k_c} as arguments and outputs the corresponding amplitude sequence $a^{n_c} = f_{ccdm}(u^{k_c}) \in \mathcal{T}_{P_{\bar{A}}}^{n_c}$. In particular, we use the same CCDM implementation for different transmission rates. In our simulations (see Sec. IX), the matcher/dematcher of our CCDM implementation performed faster than the LDPC encoder/decoder of the Communications System Toolbox of Matlab (R2014b). For binary output ($|\mathcal{A}| = 2$), an algorithm similar to [11] was proposed in [65].

VI. Bit-Metric Decoding

The receiver estimates the transmitted codeword X^{n_c} from the channel outputs Y^{n_c}. In this section, we show how this can be implemented by a bit-metric decoder.

A. Preliminaries: Binary Labeling and Decoding

Consider a coded modulation system where the input X takes values in a 2^m-ASK constellation. An optimal decoder uses the symbol-metric $p_{Y|X}$ [13, Sec. II.B] and can achieve the rate

$$R_{SMD} = \mathbb{I}(X; Y) \quad (49)$$

where SMD stands for *symbol-metric decoding* (SMD). We are interested in successfully decoding at a transmission rate close to R_{SMD} by using a *binary* decoder. Recall that our PAS scheme labels the amplitude $A = |X|$ by a length $m - 1$ binary string $\mathbf{b}(A)$ and the sign $S = \text{sign}(X)$ by one bit $b(S)$. The length m binary string

$$\mathbf{B} = B_1 B_2 \cdots B_m := b(S)\mathbf{b}(A) \quad (50)$$

with distribution $P_{\mathbf{B}}$ on $\{0, 1\}^m$ assigns to each signal point $x \in \mathcal{X}$ a label via

$$\text{label}(x) = \text{label}(\text{sign}(x))\text{label}(|x|) = b_1 b_2 \cdots b_m. \quad (51)$$

Since the labeling is one-to-one, we can also select a signal point for transmission by choosing the label $\mathbf{B}$, i.e., we use

$$X = x_{\mathbf{B}} = \{x \in \mathcal{X}: \text{label}(x) = \mathbf{B}\}. \quad (52)$$

We can interpret $\mathbf{B}$ as the channel input and the input/output relation of our channel as

$$Y = \Delta x_{\mathbf{B}} + Z. \quad (53)$$

Using the chain rule, we expand the mutual information of $\mathbf{B}$ and Y as

$$\mathbb{I}(\mathbf{B}; Y) = \sum_{i=1}^{m} \mathbb{I}(B_i; Y | B^{i-1}) \quad (54)$$

$$= \mathbb{I}(B_1; Y) + \mathbb{I}(B_2; Y | B_1)$$

$$+ \cdots + \mathbb{I}(B_m; Y | B_1 \cdots B_{m-1}). \quad (55)$$

This expansion suggests the following binary decoding:

1) Use the channel output Y to calculate an estimate $\hat{B}_1$.
2) Successively use the output Y and the estimates $\hat{B}_1 \cdots \hat{B}_{i-1}$ to calculate the next estimate $\hat{B}_i$.

This approach is called *multistage decoding* (MD). It requires *multilevel coding* (MLC) at the transmitter, i.e., on each bit-level, an individual binary code with block length n_c is used. MLC/MD was first introduced in [66] and it is discussed in detail, e.g., in [27]. To use MLC/MD, we would need to modify our PAS scheme. A simpler approach is to ignore the

estimates $\hat{B}_j$, $j \neq i$ when estimating B_i. This reduces the mutual information, which can be seen as follows

$$\mathbb{I}(\boldsymbol{B}; Y) = \sum_{i=1}^{m} \mathbb{I}(B_i; Y|B^{i-1}) \tag{56}$$

$$= \sum_{i=1}^{m} \mathbb{H}(B_i|B^{i-1}) - \mathbb{H}(B_i|YB^{i-1}) \tag{57}$$

$$\overset{(a)}{\geq} \sum_{i=1}^{m} \mathbb{H}(B_i|B^{i-1}) - \mathbb{H}(B_i|Y) \tag{58}$$

$$\overset{(b)}{=} \mathbb{H}(\boldsymbol{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y) \tag{59}$$

where (a) follows because conditioning does not increase entropy [48, Theorem 2.6.5] and where we used the chain rule for entropy in (b). The expression (59) can be achieved as follows. We jointly encode all bit-levels by a single binary code of block length mn_c. At the receiver, we use a *bit-metric decoder*. This idea was introduced in [28] and is now usually called *bit-interleaved coded modulation* (BICM) [14], [29]. Since our PAS transmitter encodes all bit-levels by a single binary code $\boldsymbol{G} = [\boldsymbol{I}_k|\boldsymbol{P}]$, our transmitter is a BICM encoder. Note that a bit-interleaver is implicitly present in the parity matrix $\boldsymbol{P}$. In Sec. VII, we discuss how $\boldsymbol{P}$ can be implemented by combining the parity matrix of an off-the-shelf code with a bit-interleaver.

B. Bit-Metric Decoding

A soft-demapper calculates for each bit-level i

$$L_i = \underbrace{\log \frac{P_{B_i}(0)}{P_{B_i}(1)}}_{\text{a-priori information}} + \underbrace{\log \frac{p_{Y|B_i}(Y|0)}{p_{Y|B_i}(Y|1)}}_{\text{channel likelihood}}. \tag{60}$$

The distribution P_{B_i} and the conditional density $p_{Y|B_i}$ in (60) can be calculated as

$$P_{B_i}(b_i) = \sum_{\boldsymbol{a} \in \{0,1\}^m : a_i = b_i} P_{\boldsymbol{B}}(\boldsymbol{a}) \tag{61}$$

$$p_{Y|B_i}(y|b_i) = \sum_{\boldsymbol{a} \in \{0,1\}^m : a_i = b_i} p_{Y|\boldsymbol{B}}(y|\boldsymbol{a}) \frac{P_{\boldsymbol{B}}(\boldsymbol{a})}{P_{B_i}(b_i)}. \tag{62}$$

By [67, Sec. 8.2.1], L_i is a sufficient statistic to estimate bit level B_i from the channel output Y, i.e., we have $\mathbb{I}(B_i; Y) = \mathbb{I}(B_i; L_i)$. A bit-metric decoder uses $L_1, L_2, \ldots, L_m$ to estimate the transmitted data and achieves the rate [12, Theorem 1]

$$\mathsf{R}_{\mathrm{BMD}} = \mathbb{H}(\boldsymbol{B}) - \sum_{i=1}^{m} \mathbb{H}(B_i|Y). \tag{63}$$

Note that (63) is equal to the expression we derived in (59).

Remark 4: If the bit levels are independent, the rate $\mathsf{R}_{\mathrm{BMD}}$ becomes the "BICM capacity" [13, Theorem 1]

$$\sum_{i=1}^{m} \mathbb{I}(B_i; Y). \tag{64}$$

TABLE II
TWO LABELINGS FOR THE AMPLITUDES OF 8-ASK AND THE RESULTING SIGNAL POINT LABELINGS

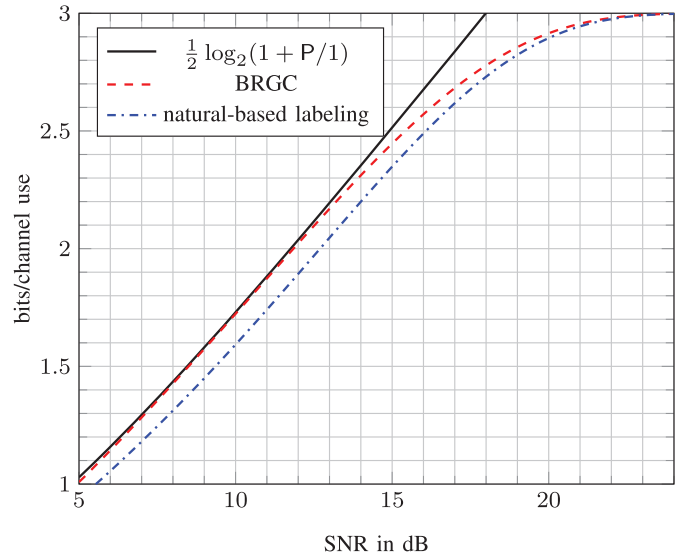| Amplitude	7	5	3	1				
natural labeling	00	01	10	11				
BRGC	00	01	11	10				
Signal point	-7	-5	-3	-1	1	3	5	7
natural-based	000	001	010	011	111	110	101	100
BRGC	000	001	011	010	110	111	101	100

Fig. 8. Comparison of $\mathsf{R}_{\mathrm{BMD}}$ for the two labelings from Table II.

Using non-uniform distributions for the independent bit-levels results in bit-shaping, which was introduced in [68] and is discussed in [12]. The authors of [69, Eq. (1)] call (64) the "pragmatic capacity".

From a coding perspective, bit-metric decoding transforms the channel (53) into m parallel binary input channels $p_{L_i|B_i}$, $i = 1, 2, \ldots, m$.

1) The bit channels $p_{L_i|B_i}$ are different for different bit-levels. This is important in Sec. VII, where we optimize the parity matrix $\boldsymbol{P}$ for bit-metric decoding.
2) By (62), the distribution of the bits B_j, $j \neq i$ influences the channel transition probabilities $p_{L_i|B_i}$ of the ith bit-channel, i.e., for each bit level, the other bit-levels act as interference.
3) The rate (63) can take on different values for different labelings of the signal points. We discuss this next.

C. Optimizing the Labeling

We choose the labeling of the $m - 1$ amplitudes; the label of the sign is already defined by construction. We evaluate $\mathsf{R}_{\mathrm{BMD}}$ for 8-ASK and two different amplitude labelings, namely the *natural labeling* and the *Binary Reflected Gray Code* (BRGC), see [70]. The two labelings are listed in Table II. We have ordered the amplitudes in descending order so that the actual signal points are labeled from left to right. The resulting labelings of the signal points are also displayed in Table II. Note that the natural labeling of the amplitudes does not lead to a natural labeling of the signal points, we therefore call it the

TABLE III
COMPARISON OF SNRs NEEDED BY SYMBOL-METRIC DECODING AND BIT-METRIC DECODING TO ACHIEVE A CERTAIN RATE

constellation	rate	SNR(R_{SMD}) [dB]	SNR(R_{BMD}) [dB]	Gap [dB]
4-ASK	1	4.8180	4.8313	0.0133
8-ASK	2	11.8425	11.8481	0.0056
16-ASK	3	18.0911	18.0951	0.0039
32-ASK	4	24.1708	24.1742	0.0034
64-ASK	5	30.2078	30.2110	0.0032

TABLE IV
VARIABLE NODE DEGREE DISTRIBUTIONS OF DVB-S2 CODES

rate	variable node degrees						
	13	12	11	4	3	2	1
2/3	4320				38880	21599	1
3/4		5400			43200	16199	1
4/5			6480		45360	12959	1
5/6	5400				48600	10799	1
9/10				6480	51840	6479	1

TABLE V
OPTIMAL BIT-LEVEL INTERLEAVER FOR UNIFORM INPUTS AND DVB-S2 CODES

constellation	code rate	bit-mapper π
8-ASK	2/3	(3, 2, 1)
16-ASK	3/4	(3, 4, 2, 1)
32-ASK	4/5	(3, 2, 5, 4, 1)
64-ASK	5/6	(4, 2, 5, 3, 6, 1)

natural-based labeling. In contrast, the BRGC labeling of the amplitudes also leads to the BRGC of the signal points. We display in Fig. 8 R_{BMD} for the natural-based labeling and the BRGC labeling. The BRGC labeling is better than the natural-based labeling and very close to the capacity $C(P)$, which is consistent with the results presented in [12]. In Table III, bit-metric decoding with BRGC is compared to symbol-metric decoding. The SNRs needed to achieve a given rate are listed.

Remark 5: The bit-metric decoding gaps in Table III show how much we can gain by using MLC/MD or by iteratively exchanging extrinsic information between the binary soft-demapper and the binary soft-decoder (BICM-ID, [71]). This gain is negligible and not worth the increased complexity for the considered scenario.

VII. BIT-MAPPER OPTIMIZATION FOR THE DVB-S2 CODES

In principle, our scheme works for any binary code with a systematic encoder and a decoder that can process the soft-output of the binary demappers. In this section, we discuss the deployment of the DVB-S2 LDPC codes.

A. LDPC Codes and Bit-Channels

LDPC codes are linear block codes with a sparse $(n - k) \times n$ check matrix H. The matrix H can be represented by a Tanner graph [72] consisting of variable nodes and check nodes. The variable node degree of the ith coded bit is given by the number of ones in the ith column of H and the degree of the jth check node is given by the number of ones in the jth row of H. The variable and check node degrees influence the decoding threshold of the LDPC code [73].

Good LDPC codes are often irregular, i.e., not all coded bits have the same variable node degree [73], [74]. At the same time, the coded bits are transmitted over different bit-channels. This suggests that the bit-mapper, which decides which coded bit is transmitted over which bit-channel, influences the performance. Bit-mapper optimization was considered, e.g., in [75]–[79]. An alternative approach is to jointly optimize the node degrees and the bit-mapper. This is done in [80] and [81]. In this work, we consider bit-mapper optimization.

B. Bit-Mapper Optimization

We use LDPC codes from the DVB-S2 standard. In Table IV, we display the variable node degree distributions of the DVB-S2 codes, e.g., the rate 2/3 code has 4320 coded bits of variable node degree 13, and it has 38880, 21599, and 1 bits

of degrees 3, 2, and 1, respectively. All codes have four different variable node degrees, which decrease from left to right. By Sec. IV-B, our PAS scheme places the uniformly distributed bits of bit-level B_1 at the end of the codeword in the systematic encoding process. For the bit-levels $B_2 \cdots B_m$, we use the following heuristic.

- A bit interleaver π_b sorts the bit stream by bit-levels, i.e.,

$$b(A_1) \ldots b(A_{n_c}) \xrightarrow{\pi_b} B_2 B_3 \cdots B_m \quad (65)$$

where B_i is a string of n_c bits of level i.

- A bit-level interleaver permutes the bit-level strings. We compactly represent the bit-level interleaver by listing the bit-levels in the order in which they should occur in the codeword, e.g., for $m = 4$, the bit-level interleaver $(4, 2, 3)$ is

$$B_2 B_3 B_4 \xrightarrow{(4,2,3)} B_4 B_2 B_3. \quad (66)$$

We keep the bit interleaver fixed and represent the complete bit-mapping by appending a 1 to the bit-level interleaver, e.g., $\pi = (4, 2, 3, 1)$ stands for the composition of π_b with bit-level interleaver $(4, 2, 3)$. We optimize the bit-level interleaver. There are $(m - 1)!$ possibilities, among which we choose the one with the best error performance. The largest considered constellation is 64-ASK, for which we need to choose among $(6 - 1)! = 120$ bit level interleavers. This optimization is still feasible. We display in Table V the optimized bit-mappers for uniform inputs and rate $(m - 1)/m$ codes.

VIII. RATE ADAPTION

A. Practical Operating Points

We use PAS at the transmitter and bit-metric decoding at the receiver. The transmission rate and the achievable rate are given by the respective

$$R = \mathbb{H}(A) + \gamma, \qquad R_{BMD} = \mathbb{H}(B) - \sum_{i=1}^{m} H(B_i | L_i) \quad (67)$$
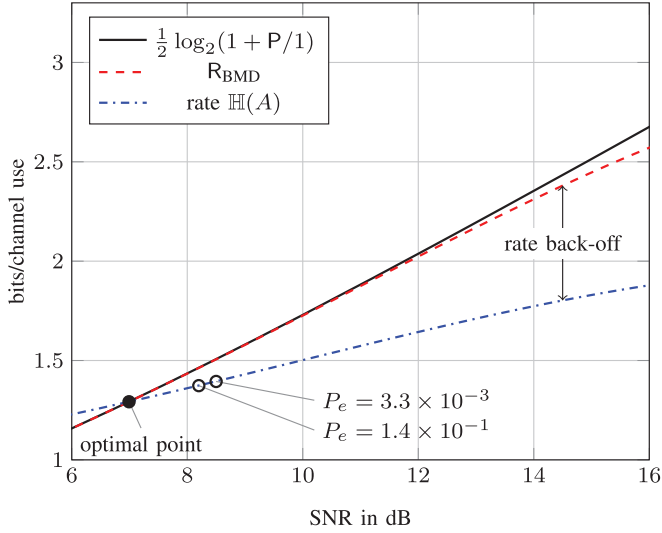
Fig. 9. Operating points for 8-ASK and rate 2/3 codes. We use the bit-level interleaver $\pi = (3, 2, 1)$, see Sec. VII-B. The optimal point is where the rate curve $R(\mathsf{P})$ crosses the achievable rate curve $\mathsf{R_{BMD}}(\mathsf{P})$. By backing off from the optimal point along the rate curve, the rate back-off (given by the distance between the rate curve and the achievable rate curve) increases. For the rate 2/3 DVB-S2 LDPC code, the FER P_e is decreased from 1.1×10^{-1} to 2.8×10^{-3} by increasing the rate back-off. SNR and rate are 8.2 dB and 1.3735 bits/channel use for $P_e = 1.1 \times 10^{-1}$ and 8.5 dB and 1.3945 bits/channel use for $P_e = 2.8 \times 10^{-3}$.

where $\boldsymbol{B} = b(S)\boldsymbol{b}(A)$. For each transmit power P, we choose the amplitude distribution P_A and the constellation scaling Δ to maximize the achievable rate. In this way, we obtain a transmission rate curve R and an achievable rate curve $\mathsf{R_{BMD}}$. In analogy to our discussion in Sec. IV-C, the optimal operating point for our scheme is where R crosses $\mathsf{R_{BMD}}$. We illustrate this in Fig. 9 for 8-ASK and code rate $c = 2/3$ ($\gamma = 0$). When we use practical codes of finite block length n_c, we must back off from the optimal operating point and tolerate a positive FER. We back off along the transmission rate curve R. As we increase the transmission power P, the *rate back-off*

$$\mathsf{R_{BMD}} - R \tag{68}$$

increases and we expect that the error probability decreases. This intuition is confirmed by the practical operating points of the 2/3 DVB-S2 LDPC code shown in Fig. 9.

B. Transmission Rate Adaption

Suppose we achieve a transmission rate R° at a FER P_e by the procedure described in Sec. VIII-A. Suppose further that this *reference operating point* is achieved by amplitude distribution P_A and constellation scaling Δ. With the same code, we now want to transmit at a rate $\tilde{R} \neq R^\circ$ and achieve the same error probability P_e.

Recall that for some optimized ν (see Sec. III-C), the input distribution $P_{X^\clubsuit}$ is given by

$$P_{X^\clubsuit}(x) = \frac{e^{-\nu x^2}}{\sum_{x' \in \mathcal{X}} e^{-\nu x'^2}}, \quad x \in \mathcal{X}. \tag{69}$$

By (16), we can write the amplitude distribution P_A of our reference point as

$$P_A(\alpha) = 2 P_{X^\clubsuit}(\alpha), \quad \alpha \in \mathcal{A}. \tag{70}$$

Now define

$$P_{A^\lambda}(\alpha) = \frac{P_A(\alpha)e^{\lambda \alpha^2}}{\sum_{\alpha' \in \mathcal{A}} P_A(\alpha')e^{\lambda \alpha'^2}}, \quad \alpha \in \mathcal{A}. \tag{71}$$

The distribution P_{A^λ} has the following properties:
- P_A and P_{A^λ} are Maxwell-Boltzmann distributions. In particular, the set of Maxwell-Boltzmann distributions is closed under the mapping $P_A \mapsto P_{A^\lambda}$ as defined by (71).
- For $\lambda = 0$, we have $P_{A^\lambda} = P_A$.
- For $\lambda \to \nu$, P_A approaches the uniform distribution on $\mathcal{A}$ whose entropy is $\log_2 |\mathcal{A}| = m - 1$.
- For $\lambda \to -\infty$, P_{A^λ} approaches the distribution that chooses the smallest amplitude with probability 1. The resulting entropy is zero.

From these properties, we see that we can use P_{A^λ} to adapt the transmission rate. The range of feasible rates is

$$\gamma \leq \mathbb{H}(A^\lambda) + \gamma \leq m - 1 + \gamma. \tag{72}$$

For $\gamma = 0$, the rate is between 0 and $m - 1$. To transmit at a feasible rate $\tilde{R}$, we proceed as follows.
- Choose the amplitude distribution such that the desired transmission rate is achieved, i.e., choose

$$P_{\tilde{A}} = P_{A^\lambda} : \mathbb{H}(A^\lambda) + \gamma = \tilde{R}. \tag{73}$$

- Choose the constellation scaling $\tilde{\Delta}$ such that the resulting error probability $\tilde{P}_e$ is the same as for the reference point, i.e., choose

$$\tilde{\Delta} : \tilde{P}_e = P_e. \tag{74}$$

C. Adaption for Robust Codes

The procedure described in the previous section requires to search for the right constellation scaling $\tilde{\Delta}$ by repeatedly performing Monte Carlo simulations, which is time consuming. We therefore seek for a simpler method to find $\tilde{\Delta}$ by exploring the relationship between P_e and the rate back-off. We start by discussing two special cases.

Example 2: (Rate Back-off for Uniform Binary Input). For a uniformly distributed binary input B and a rate c code, the transmission rate is $R = c$ and the achievable rate is $R^* = \mathbb{I}(B; Y)$. The rate back-off becomes

$$R^* - R = \mathbb{I}(B; Y) - c. \tag{75}$$

In [82], it was observed for practical LDPC codes that P_e is approximately determined by the mutual information $\mathbb{I}(B; Y)$. Since c in (75) is a property of the considered code and independent of the channel, this is equivalent to saying that P_e is approximately determined by the rate back-off.

Example 3: (Rate Back-off for BICM with Uniform Input). Consider a BICM scheme with a binary code of rate c and a constellation with 2^m signal points. Let $B_1, B_2, \ldots, B_m$ be the corresponding bit-levels. The achievable rate is $R^* = \sum_{i=1}^{m} \mathbb{I}(B_i; Y)$ and for uniform inputs, the transmission rate is $R = c \cdot m$. The rate back-off becomes

$$R^* - R = \sum_{i=1}^{m} \mathbb{I}(B_i; Y) - c \cdot m. \tag{76}$$

In [83], it was observed that the achievable rate R^* approximately determines P_e. A similar observation was made in [84]. Since $c \cdot m$ in (76) is a property of the BICM scheme and independent of the channel, this is equivalent to saying that the rate back-off approximately determines P_e.

We now apply the observations from Example 2 & 3 to our scheme. We say a code is *robust*, if P_e is approximately determined by the rate back-off $R_{BMD} - R$. Suppose for P_A and Δ of our reference point, the achievable rate and the transmission rate evaluate to R_{BMD}° and R°, respectively. If the considered code is robust, we can adapt to a transmission rate $\tilde{R}$ as follows.

• Choose the amplitude distribution $P_{\tilde{A}}$ according to (73).
• Choose the constellation scaling $\tilde{\Delta}$ such that the resulting rate back-off is equal to the rate back-off of the reference point, i.e.,

$$\tilde{\Delta}: R_{BMD}(\tilde{\Delta}, P_{\tilde{A}}) - \tilde{R} = R_{BMD}^\circ - R^\circ. \tag{77}$$

In Sec. IX, we apply this strategy to the DVB-S2 LDPC codes and we assess their robustness by evaluating the resulting FERs.

Remark 6: For our scheme, the rate back-off can be written as

$$R_{BMD} - R = 1 - \gamma - \sum_{i=1}^{m} \mathbb{H}(B_i | L_i). \tag{78}$$

Since the term $1 - \gamma$ is independent of the channel and the input distribution, adjusting the rate back-off corresponds to adjusting the sum of the conditional entropies $\sum_{i=1}^{m} \mathbb{H}(B_i | L_i)$. This property is explored for code design in [81, Sec. III.B].

Remark 7: The authors of [85], [86] investigate the analysis and design of universal LDPC codes, which provably perform well on different channels.

IX. NUMERICAL RESULTS

We assess the performance of our scheme by Monte Carlo simulation. For each block, k_c uniformly distributed data bits are transmitted in n_c channel uses. γn_c of the data bits $U^{k_c} = U_1 U_2 \cdots U_{k_c}$ are used for bit-level 1 (see Fig. 5). The remaining $k_c - \gamma n_c$ data bits are transformed by a CCDM matcher (see Sec. V) into a sequence of n_c amplitudes. Encoding is then done according to Fig. 5. At the receiver, bit-metric decoding is performed (see Sec. VI) and the amplitude estimates are transformed back into $k_c - \gamma n_c$ data bit estimates by a CCDM dematcher, see [11]. The dematcher output together with the other γn_c data bit estimates form the data estimate $\hat{U}^{k_c} = \hat{U}_1 \hat{U}_2 \cdots \hat{U}_{k_c}$. We estimate the end-to-end FER $\Pr\{\hat{U}^{k_c} \neq U^{k_c}\}$.

TABLE VI
64-ASK, 9/10 DVB-S2, $\pi = (4, 2, 5, 3, 6, 1)$

Rate	SNR [dB]	Gap [dB]	FER	95% CI
5.09	31.80	1.15	$4.1 \cdot 10^{-3}$	$\pm 2.1 \cdot 10^{-3}$
4.98	31.09	1.09	$8.0 \cdot 10^{-3}$	$\pm 4.0 \cdot 10^{-3}$
4.89	30.46	1.05	$6.8 \cdot 10^{-3}$	$\pm 3.4 \cdot 10^{-3}$
4.79	29.84	1.03	$7.3 \cdot 10^{-3}$	$\pm 3.7 \cdot 10^{-3}$
4.69	29.23	1.02	$6.4 \cdot 10^{-3}$	$\pm 3.2 \cdot 10^{-3}$
4.59	28.62	1.02	$3.8 \cdot 10^{-3}$	$\pm 1.9 \cdot 10^{-3}$
4.49	28.01	1.01	$6.2 \cdot 10^{-3}$	$\pm 3.1 \cdot 10^{-3}$
4.39	27.41	1.01	$7.6 \cdot 10^{-3}$	$\pm 3.8 \cdot 10^{-3}$
4.29	26.80	1.00	$9.6 \cdot 10^{-3}$	$\pm 4.8 \cdot 10^{-3}$
4.19	26.21	1.00	$1.0 \cdot 10^{-2}$	$\pm 5.1 \cdot 10^{-3}$
4.09	25.61	1.00	$1.0 \cdot 10^{-2}$	$\pm 5.1 \cdot 10^{-3}$
3.99	25.00	0.99	$1.0 \cdot 10^{-2}$	$\pm 5.0 \cdot 10^{-3}$

TABLE VII
32-ASK, 5/6 DVB-S2, $\pi = (4, 5, 2, 3, 1)$

Rate	SNR [dB]	Gap [dB]	FER	95% CI
3.62	22.60	0.82	$2.1 \cdot 10^{-3}$	$\pm 1.1 \cdot 10^{-3}$
3.99	25.09	1.07	$3.3 \cdot 10^{-3}$	$\pm 1.7 \cdot 10^{-3}$
3.89	24.36	0.94	$1.8 \cdot 10^{-3}$	$\pm 9.0 \cdot 10^{-4}$
3.79	23.69	0.87	$2.8 \cdot 10^{-3}$	$\pm 1.4 \cdot 10^{-3}$
3.69	23.05	0.84	$1.5 \cdot 10^{-3}$	$\pm 8.0 \cdot 10^{-4}$
3.59	22.42	0.82	$1.6 \cdot 10^{-3}$	$\pm 8.0 \cdot 10^{-4}$
3.49	21.80	0.81	$1.5 \cdot 10^{-3}$	$\pm 8.0 \cdot 10^{-4}$
3.39	21.19	0.80	$1.3 \cdot 10^{-3}$	$\pm 6.0 \cdot 10^{-4}$
3.29	20.58	0.80	$8.0 \cdot 10^{-4}$	$\pm 4.0 \cdot 10^{-4}$
3.19	19.98	0.80	$7.0 \cdot 10^{-4}$	$\pm 4.0 \cdot 10^{-4}$
3.10	19.37	0.80	$5.0 \cdot 10^{-4}$	$\pm 3.0 \cdot 10^{-4}$
3.00	18.76	0.80	$7.0 \cdot 10^{-4}$	$\pm 4.0 \cdot 10^{-4}$

For each FER estimate, we also provide the corresponding 95% *confidence interval* (CI). The spectral efficiency is $R = \frac{k_c}{n_c}$. We use the DVB-S2 LDPC codes and we decode with 100 iterations. For each considered (constellation, code rate) mode, we first determine the practical operating point with FER $\approx 1 \times 10^{-3}$ following the procedure described in Sec. VIII-A. We also optimize the bit-mapper using the heuristic proposed in Sec. VII-B. We then use the rate adaption (77) to operate the same (constellation, code rate) mode over a range of SNRs and spectral efficiencies. We use the same bit-mapper for all operating points of the same (constellation, code rate) mode. The results are displayed in Table VI–X. In the first line of each table, the practical operating point is displayed and separated from the rest of the displayed values by a horizontal line. For 8-ASK, we also display the resulting operating points in Fig. 10.

1) Spectral Efficiency: In Table VI—X, we display the SNR gap of our scheme to capacity $C(P) = \frac{1}{2} \log_2(1 + P/1)$. Over a range of 1 bit/s/Hz to 5 bit/s/Hz, our scheme operates within 1.1 dB of capacity. The rate 5/6 code with 16-ASK has a gap of only 0.61 to 0.69 dB, while the rate 9/10 code with 64-ASK has a gap of 1 dB. This indicates that the spectral efficiency can be improved further by using optimized codes. The work [81] confirms this.

2) Robustness: Recall that the listed operating points were found by (77) with the practical operating point from

TABLE VIII
16-ASK, 5/6 DVB-S2, $\pi = (4, 3, 2, 1)$

Rate	SNR [dB]	Gap [dB]	FER	95% CI
2.96	18.40	0.67	$2.0 \cdot 10^{-2}$	$\pm 9.8 \cdot 10^{-3}$
3.00	18.66	0.69	$3.0 \cdot 10^{-2}$	$\pm 1.4 \cdot 10^{-2}$
2.90	18.01	0.65	$2.3 \cdot 10^{-2}$	$\pm 1.2 \cdot 10^{-2}$
2.80	17.38	0.63	$2.0 \cdot 10^{-2}$	$\pm 9.8 \cdot 10^{-3}$
2.70	16.75	0.62	$1.6 \cdot 10^{-2}$	$\pm 7.8 \cdot 10^{-3}$
2.60	16.13	0.61	$1.3 \cdot 10^{-2}$	$\pm 6.7 \cdot 10^{-3}$
2.50	15.51	0.61	$7.2 \cdot 10^{-3}$	$\pm 3.6 \cdot 10^{-3}$
2.40	14.89	0.61	$3.5 \cdot 10^{-3}$	$\pm 1.8 \cdot 10^{-3}$
2.30	14.27	0.62	$2.5 \cdot 10^{-3}$	$\pm 1.3 \cdot 10^{-3}$
2.20	13.64	0.62	$1.9 \cdot 10^{-3}$	$\pm 9.0 \cdot 10^{-4}$
2.10	13.01	0.62	$1.3 \cdot 10^{-3}$	$\pm 7.0 \cdot 10^{-4}$
2.00	12.37	0.63	$5.0 \cdot 10^{-4}$	$\pm 2.0 \cdot 10^{-4}$

TABLE IX
8-ASK, 3/4 DVB-S2, $\pi = (3, 2, 1)$

Rate	SNR [dB]	Gap [dB]	FER	95% CI
1.85	11.45	0.63	$1.5 \cdot 10^{-3}$	$\pm 8.0 \cdot 10^{-4}$
2.00	12.44	0.69	$3.8 \cdot 10^{-3}$	$\pm 1.9 \cdot 10^{-3}$
1.90	11.75	0.64	$2.1 \cdot 10^{-3}$	$\pm 1.1 \cdot 10^{-3}$
1.80	11.08	0.62	$1.2 \cdot 10^{-3}$	$\pm 6.0 \cdot 10^{-4}$
1.70	10.41	0.62	$2.3 \cdot 10^{-3}$	$\pm 1.2 \cdot 10^{-3}$
1.60	9.75	0.62	$1.2 \cdot 10^{-3}$	$\pm 6.0 \cdot 10^{-4}$
1.50	9.07	0.63	$1.5 \cdot 10^{-3}$	$\pm 7.0 \cdot 10^{-4}$
1.40	8.39	0.64	$1.2 \cdot 10^{-3}$	$\pm 6.0 \cdot 10^{-4}$
1.30	7.69	0.65	$2.2 \cdot 10^{-3}$	$\pm 1.1 \cdot 10^{-3}$

TABLE X
4-ASK, 2/3 DVB-S2, $\pi = (2, 1)$

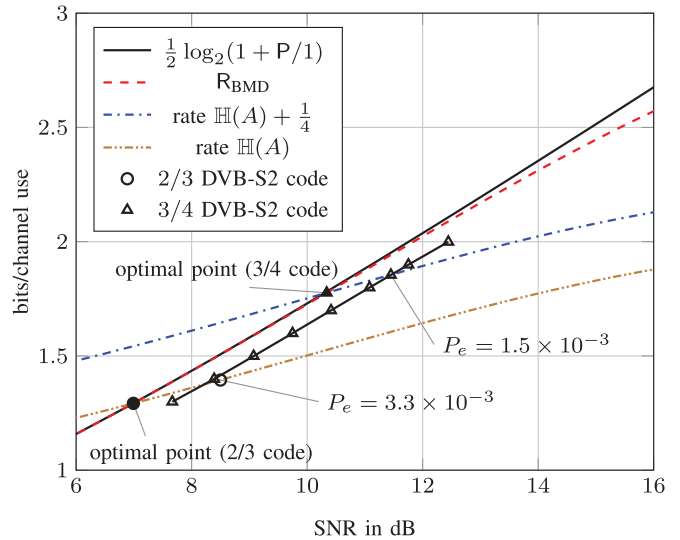| Rate	SNR [dB]	Gap [dB]	FER	95% CI
1.13	6.70	0.90	$5.2 \cdot 10^{-3}$	$\pm 2.6 \cdot 10^{-3}$
1.20	7.26	0.94	$2.6 \cdot 10^{-3}$	$\pm 1.3 \cdot 10^{-3}$
1.10	6.45	0.90	$6.4 \cdot 10^{-3}$	$\pm 3.3 \cdot 10^{-3}$
1.00	5.66	0.90	$1.4 \cdot 10^{-2}$	$\pm 6.8 \cdot 10^{-3}$

Fig. 10. Transmission rate adaption for 8-ASK. The rate 3/4 DVB-S2 code is used. The triangles mark the operating points calculated under the assumption that the code is robust, see Sec VIII-C. The reference operating point is the triangle on the rate $\mathbb{H}(A) + \frac{1}{4}$ curve. In Table IX, we display the FER of the adapted operating points. For comparison, we display an operating point of the 2/3 DVB-S2 code.

the smaller constellations 4, 8 and 16-ASK, the effective rate is equal to $\mathbb{H}(A)$ with a precision of two decimal places. For the larger constellations 32 and 64-ASK, we observe a slight decrease of the effective rate, e.g., instead of the target rate 5.00 we observe an effective rate $R = 4.98$. See also [11] for a discussion of this phenomenon. We calculate the SNR gap with respect to the effective rate.

4) Error Floor: The DVB-S2 codes are designed to get the FER down to a reasonable value and an outer code then lowers the error probability further [1]. Since the CCDM dematcher performs a non-linear transformation, one corrupted symbol at the dematcher input can lead to various corrupted bit-errors at the dematcher output, so in-block error propagation can occur. In our scheme, an outer encoder should therefore be placed between the matcher and the inner encoder and the outer decoder should be placed between the inner decoder and the dematcher. To preserve the amplitude distribution imposed by the matcher, a systematic outer encoder has to be used.

X. CONCLUSIONS

We proposed a practical rate-matched coded modulation scheme that adapts its transmission rate to the SNR. At the transmitter, a distribution matcher is concatenated with a systematic encoder of a binary LDPC code. At the receiver, bit-metric decoding is used. No iterative demapping is required. The reported numerical results show that on the complex baseband AWGN channel, any spectral efficiency between 2–10 bit/s/Hz can be achieved within 1.1 dB of capacity $\log_2(1 + \mathsf{P}/1)$ by using the off-the-shelf DVB-S2 LDPC codes of rate 2/3, 3/4, 4/5 and 9/10 together with QAM constellations with up to 4096 signal points. Future work should investigate rate-matched coded modulation for short block lengths and for multiple-input multiple-output channels.

Sec. VIII-A as reference point. The results in Table VI—X show that the DVB-S2 codes are robust in the sense of Sec. VIII-C, since over the whole considered range, the resulting FER is within the waterfall region. Some configurations show a "more robust" behavior than others, e.g., the FER of the rate 3/4 code with 8-ASK (Table IX) is almost the same for the spectral efficiencies from 1.3 to 2.0 bits/s/Hz, while for the rate 5/6 code with 16-ASK, the FER changes by almost two orders of magnitude. Because of limited computational resources, we were not able to simulate FERs below 10^{-4}. To keep the FERs for 16-ASK above 10^{-4}, we had to increase the target FER of the practical operating point to $\approx 10^{-2}$. In principle, the operating points of 16-ASK that result in a too low (too high) FER could be adjusted by decreasing (increasing) the transmission power.

3) Rate: Since all considered codes have the same block length of 64800 bits, the number of channel uses n_c gets smaller for larger constellations. As discussed in Sec. V, the CCDM matcher that we use to emulate the amplitude source P_A has a rate close to $\mathbb{H}(A)$ when the output length n_c is large. For

APPENDIX

Let D_1 and D_2 be two independent binary random variables with distributions P and Q, respectively, and define $R = D_1 \oplus D_2$. Without loss of generality, assume that

$$P(0) \leq P(1), \quad Q(0) \leq Q(1). \tag{79}$$

Since $P_R(1) = 1 - P_R(0)$, the distribution P_R is more uniform than P and Q if

$$\min\{P(1), Q(1)\} \geq P_R(0) \geq \max\{P(0), Q(0)\}. \tag{80}$$

By $P_R(0) = P(0)Q(0) + P(1)Q(1)$, we have

$$P_R(0) \geq P(0)Q(0) + P(1)Q(0) = Q(0) \tag{81}$$

$$P_R(0) \geq P(0)Q(0) + P(0)Q(1) = P(0) \tag{82}$$

$$P_R(0) \leq P(0)Q(1) + P(1)Q(1) = Q(1) \tag{83}$$

$$P_R(0) \leq P(1)Q(0) + P(1)Q(1) = P(1). \tag{84}$$

These four inequalities imply (80).

REFERENCES

[1] *Digital Video Broadcasting (DVB); 2nd Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications (DVB-S2)*, European Telecommun. Standards Inst. (ETSI) Standard EN 302 307, Rev. 1.2.1, 2009.

[2] D. Raphaeli and A. Gurevitz, "Constellation shaping for pragmatic turbo-coded modulation with high spectral efficiency," *IEEE Trans. Commun.*, vol. 52, no. 3, pp. 341–345, Mar. 2004.

[3] M. Yankov, S. Forchhammer, K. J. Larsen, and L. P. Christensen, "Rate-adaptive constellation shaping for near-capacity achieving turbo coded BICM," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2014, pp. 2112–2117.

[4] R. De Gaudenzi, A. Guillén i Fàbregas, A. M. Vicente, and B. Ponticelli, "APSK coded modulation schemes for nonlinear satellite channels with high power and spectral efficiency," in *Proc. AIAA Int. Commun. Satell. Syst. Conf. (ICSSC)*, 2002, pp. 1–11.

[5] *Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications; Part 2: DVB-S2 Extensions (DVB-S2X)*, Eur. Telecommun. Standards Inst., (ETSI) Std. EN 302 307-2, Rev. 1.1.1, 2014.

[6] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[7] J. Li and K. R. Narayanan, "Rate-compatible low density parity check codes for capacity-approaching ARQ schemes in packet data communications," in *Proc. Int. Conf. Commun. Internet Inf. Technol. (CIIT)*, 2002, pp. 201–206.

[8] J. Ha, J. Kim, and S. W. McLaughlin, "Rate-compatible puncturing of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2824–2836, Nov. 2004.

[9] T. V. Nguyen, A. Nosratinia, and D. Divsalar, "The design of rate-compatible protograph LDPC codes," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2841–2850, Oct. 2012.

[10] T.-Y. Chen, K. Vakilinia, D. Divsalar, and R. D. Wesel, "Protograph-based raptor-like LDPC codes," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1522–1532, May 2015.

[11] P. Schulte and G. Böcherer, "Constant composition distribution matching," *IEEE Trans. Inf. Theory*, 2015, to be published.

[12] G. Böcherer, "Achievable rates for shaped bit-metric decoding," arXiv preprint, 2015 [Online]. Available: http://arxiv.org/abs/1410.8075

[13] 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, Jun. 2009.

[14] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 1–2, pp. 1–153, 2008.

[15] G. Böcherer, "Probabilistic signal shaping for bit-metric decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014, pp. 431–435.

[16] G. D. Forney, R. G. Gallager, G. Lang, F. Longstaff, and S. Qureshi, "Efficient modulation for band-limited channels," *IEEE J. Sel. Areas Commun.*, vol. 2, no. 5, pp. 632–647, Sep. 1984.

[17] C. Ling and J.-C. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[18] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," in *Proc. Allerton Conf. Commun. Control Comput.*, 2014, pp. 789–796.

[19] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

[20] F. Schreckenbach and P. Henkel, "Signal shaping using non-unique symbol mappings," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2005, pp. 1–10.

[21] G. Böcherer, "Optimal non-uniform mapping for probabilistic shaping," in *Proc. Int. ITG Conf. Source Channel Coding (SCC)*, 2013, pp. 1–6.

[22] G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," arXiv preprint, 2015, arxiv number 1307.6843.

[23] G. D. Forney, "Trellis shaping," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 281–300, Mar. 1992.

[24] S. A. Tretter, *Constellation Shaping, Nonlinear Precoding, and Trellis Coding for Voiceband Telephone Channel Modems with Emphasis on ITU-T Recommendation V. 34*. Norwell, MA, USA: Kluwer, 2002.

[25] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. Hoboken, NJ, USA: Wiley, 2002.

[26] R. Fischer, J. Huber, and U. Wachsmann, "On the combination of multi-level coding and signal shaping," in *Proc. Int. ITG Conf. Source Channel Coding (SCC)*, 1998, pp. 1–6.

[27] 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.

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

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

[30] B. P. Smith and F. R. Kschischang, "A pragmatic coded modulation scheme for high-spectral-efficiency fiber-optic communications," *J. Lightw. Technol.*, vol. 30, no. 13, pp. 2047–2053, 2012.

[31] S. Kaimalettu, A. Thangaraj, M. Bloch, and S. McLaughlin, "Constellation shaping using LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2007, pp. 2366–2370.

[32] A. K. Khandani and P. Kabal, "Shaping multidimensional signal spaces. I. Optimum shaping, shell mapping," *IEEE Trans. Inf. Theory*, vol. 39, no. 6, pp. 1799–1808, Nov. 1993.

[33] F. R. Kschischang and S. Pasupathy, "Optimal shaping properties of the truncated polydisc," *IEEE Trans. Inf. Theory*, vol. 40, no. 3, pp. 892–903, May 1994.

[34] ITU-T Recommendation V.34. (1998, Feb.). *A Modem Operating at Data Signalling Rates of Up to 33 600 bit/s for Use on the General Switched Telephone Network and on Leased Point-to-Point 2-Wire Telephone-type Circuits* [Online]. Available: http://www.itu.int/rec/T-REC-V.34-199802-I

[35] L. Duan, B. Rimoldi, and R. Urbanke, "Approaching the AWGN channel capacity without active shaping," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 1997, p. 374.

[36] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.

[37] H. S. Cronie, "Signal shaping for bit-interleaved coded modulation on the AWGN channel," *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3428–3435, Dec. 2010.

[38] S. Y. Le Goff, B. S. Sharif, and S. A. Jimaa, "Bit-interleaved turbo-coded modulation using shaping coding," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 246–248, Mar. 2005.

[39] 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, Jun. 2007.

[40] B. K. Khoo, S. Y. Le Goff, B. S. Sharif, and C. C. Tsimenidis, "Bit-interleaved coded modulation with iterative decoding using constellation shaping," *IEEE Trans. Commun.*, vol. 54, no. 9, pp. 1517–1520, Sep. 2006.

[41] M. C. Valenti and X. Xiang, "Constellation shaping for bit-interleaved LDPC coded APSK," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2960–2970, Oct. 2012.

[42] G. Böcherer and R. Mathar, "Operating LDPC codes with zero shaping gap," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2011, pp. 330–334.

[43] G. Böcherer, "Capacity-achieving probabilistic shaping for noisy and noiseless channels," Ph.D. dissertation, RWTH Aachen Univ., Aachen, Germany, 2012.

[44] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Proc. Data Compression Conf. (DCC)*, 2011, pp. 23–32.

[45] W. Bliss, "Circuitry for performing error correction calculations on baseband encoded data to eliminate error propagation," *IBM Tech. Discl. Bull.*, vol. 23, pp. 4633–4634, 1981.

[46] M. Blaum *et al.*, "High-rate modulation codes for reverse concatenation," *IEEE Trans. Magn.*, vol. 43, no. 2, pp. 740–743, Feb. 2007.

[47] F. R. Kschischang and S. Pasupathy, "Optimal nonuniform signaling for Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 913–929, May 1993.

[48] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.

[49] L. Ozarow and A. Wyner, "On the capacity of the Gaussian channel with a finite number of input levels," *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1426–1428, Nov. 1990.

[50] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[51] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[52] D. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.

[53] E. A. Ratzer, "Error-correction on non-standard communication channels," Ph.D. dissertation, Univ. Cambridge, Cambridge, U.K., 2003.

[54] R. G. Gallager, *Principles of Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[55] G. Ungerböck, "Huffman shaping," in *Codes, Graphs, and Systems*, R. Blahut, and R. Koetter, Eds. New York, NY, USA: Springer, 2002, ch. 17, pp. 299–313.

[56] N. Cai, S.-W. Ho, and R. Yeung, "Probabilistic capacity and optimal coding for asynchronous channel," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2007, pp. 54–59.

[57] R. A. Amjad and G. Böcherer, "Fixed-to-variable length distribution matching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 1511–1515.

[58] S. Baur and G. Böcherer, "Arithmetic distribution matching," in *Proc. Int. ITG Conf. Source Channel Coding (SCC)*, 2015, pp. 1–6.

[59] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[60] I. Csiszár and P. C. Shields, "Information theory and statistics: A tutorial," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[61] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K: Cambridge Univ. Press, 2011.

[62] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[63] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[64] *A Fixed-to-Fixed Length Distribution Matcher in C/MATLAB*, 2015 [Online]. Available: http://beam.to/ccdm

[65] T. V. Ramabadran, "A coding scheme for m-out-of-n codes," *IEEE Trans. Commun.*, vol. 38, no. 8, pp. 1156–1163, Aug. 1990.

[66] 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.

[67] R. G. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

[68] A. G. i Fàbregas and A. Martinez, "Bit-interleaved coded modulation with shaping," in *Proc. IEEE Inf. Theory Workshop (ITW)*, 2010, pp. 1–5.

[69] F. Kayhan and G. Montorsi, "Constellation design for transmission over nonlinear satellite channels," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, 2012, pp. 3401–3406.

[70] F. Gray, "Pulse code communication," U.S. Patent 2 632 058, 1953.

[71] X. Li and J. A. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, Nov. 1997.

[72] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inf. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[73] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[74] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, "Improved low-density parity-check codes using irregular graphs," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 585–598, Feb. 2001.

[75] Y. Li and W. E. Ryan, "Bit-reliability mapping in LDPC-coded modulation systems," *IEEE Commun. Lett.*, vol. 9, no. 1, pp. 1–3, Jan. 2005.

[76] J. Lei and W. Gao, "Matching graph connectivity of LDPC codes to high-order modulation by bit interleaving," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2008, pp. 1059–1064.

[77] L. Gong *et al.*, "Improve the performance of LDPC coded QAM by selective bit mapping in terrestrial broadcasting system," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 263–269, Jun. 2011.

[78] T. Cheng, K. Peng, J. Song, and K. Yan, "EXIT-aided bit mapping design for LDPC coded modulation with APSK constellations," *IEEE Commun. Lett.*, vol. 16, no. 6, pp. 777–780, Jun. 2012.

[79] C. Häger, A. Graell i Amat, A. Alvarado, F. Brännström, and E. Agrell, "Optimized bit mappings for spatially coupled LDPC codes over parallel binary erasure channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2064–2069.

[80] L. Zhang and F. Kschischang, "Multi-edge-type low-density parity-check codes for bandwidth-efficient modulation," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 43–52, Jan. 2013.

[81] F. Steiner, G. Böcherer, and G. Liva, "Protograph-based LDPC code design for bit-metric decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, pp. 1089–1093.

[82] M. Franceschini, G. Ferrari, and R. Raheli, "Does the performance of LDPC codes depend on the channel?" *IEEE Trans. Commun.*, vol. 54, no. 12, pp. 2129–2132, Dec. 2006.

[83] A. Alvarado, E. Agrell, D. Lavery, R. Maher, and P. Bayvel, "Replacing the soft-decision FEC limit paradigm in the design of optical communication systems," *J. Lightw. Technol.*, vol. 33, no. 20, pp. 4338–4352, 2015.

[84] A. Leven, F. Vacondio, L. Schmalen, S. ten Brink, and W. Idler, "Estimation of soft FEC performance in optical transmission experiments," *IEEE Photon. Technol. Lett.*, vol. 20, no. 23, pp. 1547–1549, Oct. 2011.

[85] I. Sason, "On universal properties of capacity-approaching LDPC code ensembles," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 2956–2990, Jul. 2009.

[86] I. Sason and B. Shuval, "On universal LDPC code ensembles over memoryless symmetric channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5182–5202, Aug. 2011.

Georg Böcherer (S'05–M'13) was born in Freiburg im Breisgau, Germany. He received the M.Sc. degree in electrical engineering and information technology from the ETH Zürich, Switzerland, in 2007, and the Ph.D. degree from the RWTH Aachen University, Aachen, Germany, in 2012. He is currently a Senior Researcher and Lecturer with the Institute for Communications Engineering, Technische Universität München, München, Germany. His research interests include coding, modulation, probabilistic shaping for optical, wireless, and wired communications.

Fabian Steiner (S'14) was born in Prien am Chiemsee, Germany. He received the B.Sc. degree and M.Sc. degree (with high distinction) in electrical engineering from the Technische Universität München (TUM), München, Germany, in 2011 and 2014, respectively. He is now working toward the Ph.D. degree in a joint research project of the Fachgebiet Methoden der Signalverarbeitung, TUM, and UC Irvine, CA, USA, with the support of the Institute for Advanced Study, TUM. His research interests include coding, modulation, and multiuser massive MIMO systems.

Patrick Schulte received the B.Sc. and M.Sc. degrees in electrical engineering from the Technische Universität München (TUM), München, Germany, in 2012 and 2014, respectively. He is currently pursuing the Ph.D. degree at the Institute for Communications Engineering, TUM. His research interests include distribution matching and coding.